

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:41:49 ; Search time 47.1111 Seconds  
(without alignments)  
1433.395 Million cell updates/sec

Title: US-09-887-784-64A  
Perfect score: 1273  
Sequence: 1 MVSGEELFTGVVPILVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_23Jan04: \*  
1: Geneseqp1980s: \*  
2: Geneseqp1990s: \*  
3: Geneseqp2000s: \*  
4: Geneseqp2001s: \*  
5: Geneseqp2002s: \*  
6: Geneseqp2003as: \*  
7: Geneseqp2003bs: \*  
8: Geneseqp2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1269	99.7	239	5	Aae17518 Enhanced
2	1269	99.7	363	6	ABR40352 Human aml
3	1269	99.7	893	4	AAG65781 Amino aci
4	1269	99.7	1132	4	AAG65782 Amino aci
5	1261	99.1	239	5	Aae17517 Enhanced
6	1258	98.8	239	3	AAB22882 Enhanced
7	1258	98.8	239	3	AAY54349 Amino aci
8	1258	98.8	239	3	AAY79584 EGFP sign
9	1258	98.8	239	4	AAB50804 Jellyfish
10	1258	98.8	239	4	AAB85900 A. victor
11	1258	98.8	239	4	AAB31171 Amino aci
12	1258	98.8	239	5	AAG66198 A. victor
13	1258	98.8	239	5	ABG94444 Protease
14	1258	98.8	239	5	Aae14599 Aequorea
15	1258	98.8	239	6	Aae34958 Aequorea
16	1258	98.8	239	6	AAG79829 Green flu
17	1258	98.8	239	6	ABR83616 Green flu
18	1258	98.8	239	6	ADA38074 Aequorea
19	1258	98.8	239	7	ABU63204 Aequorea
20	1258	98.8	239	7	ADCI18358 EGFP (enh
21	1258	98.8	239	7	ABW00914 Aequorea
22	1258	98.8	239	7	ADe28570 Enhanced
23	1258	98.8	246	7	ABW79011 Enhanced
24	1258	98.8	248	5	AAG68319 Jellyfish
25	1258	98.8	259	5	Aau99804 Biomembra

26	1258	98.8	265	2	AAW97451 Wild-type
27	1258	98.8	268	5	AAU99803 Biomembra
28	1258	98.8	270	5	AAU99802 Biomembra
29	1258	98.8	272	5	AAU99800 Biomembra
30	1258	98.8	273	5	AAU99801 Biomembra
31	1258	98.8	280	5	AAU99807 Biomembra
32	1258	98.8	281	3	AAU50142 Green flu
33	1258	98.8	281	3	AAE24252 EGFP-MODC
34	1258	98.8	281	5	AAU10888 EGFP-MODC
35	1258	98.8	286	7	ADe28562 EGFP/ hum
36	1258	98.8	289	7	ADe28564 EGFP/ hum
37	1258	98.8	290	7	ADe28568 EGFP/ hum
38	1258	98.8	290	7	ADe28566 EGFP/ hum
39	1258	98.8	294	3	AAE22860 GFP-DEVD-
40	1258	98.8	294	3	AAE24252 EGFP-MODC
41	1258	98.8	294	5	ABG94422 Recombina
42	1258	98.8	308	2	AAV42181 EGFP/DRM
43	1258	98.8	320	6	ABR83620 HUB1-GFP
44	1258	98.8	323	3	AAV54359 GFP mutan
45	1258	98.8	323	6	ABR83621 HUB1-GFP

ALIGNMENTS

RESULT 1  
AAE17518  
ID AAE17518 standard; protein; 239 AA.  
XX  
AC AAE17518;  
XX  
DT 22-APR-2002 (first entry)  
XX  
DE Enhanced F64L-E222G jellyfish green fluorescent protein mutant.  
XX  
KW Jellyfish; green fluorescent protein; GFP; protein redistribution;  
KW cellular function; genetic reporter; mutant; Stoke's shift; muten.  
XX  
OS Aequorea victoria.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT Misc-difference 65 /note= "Wild type Phe substituted with Leu; This  
FT corresponds to position 64 in the wild type protein"  
FT Misc-difference 223  
FT /note= "Wild type Glu substituted with Gly; This  
FT corresponds to position 222 in the wild type protein"

Novel fluorescent protein in vitro assay for measuring protein kinase activity or dephosphorylation activity, or for measuring protein redistribution, has a green fluorescent protein with F64L and E222G mutation.  
Claim 9; Page 37; 41pp; English.

XX The invention relates to a fluorescent protein derived from green  
CC fluorescent protein (GFP) or its analogue. The GFP containing mutations  
CC at F64L and E222G has a bigger compared to other GFP's making it very  
CC suitable for high throughput screening due to better resolution. The  
CC fluorescent protein is useful in *in vitro* assays for measuring protein  
CC kinase activity or dephosphorylation activity, or for measuring protein  
CC redistribution. The fluorescent protein is useful in studying cellular  
CC functions in living cells; as protein tags in transgenic animals, living  
CC and fixed cells; organelle tags, secretion marker and genetic reporter.  
CC The fluorescent protein is also useful as a cell or organelle integrity  
CC marker, a marker for changes in cell morphology, as transfection marker,  
CC and as a marker to be used in combination with fluorescence activated  
CC cell sorting (FACS). The novel proteins can also be used as reporters to  
CC monitor live or dead biomass of organisms, such as fungi. The fluorescent  
CC protein is also useful as markers in transcriptional and translational  
CC fusions for performing transposon vector mutagenesis and as a reporter  
CC for bacterial detection. Transposons encoding the fluorescent protein are  
CC useful for screening promoters and for tagging plasmids and chromosomes.  
CC The fluorescent protein engineered into the genome of a phase is useful  
CC for designing diagnostic tool. The present sequence is a DNA encoding  
CC enhanced F64L-E222G jellyfish green fluorescent protein (GFP) mutant  
XX  
XX  
SQ Sequence 239 AA;

Query Match 99.7%; Score 1269; DB 5; Length 239;  
Best Local Similarity 99.6%; Pred. No. 4.3e-124;  
Matches 236; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKGELFTGVWPILVELDGVNGHKFSVSGEGDGTGKLTLCFTTGKLPVDPWPT 60  
Db 1 MVSKGELFTGVWPILVELDGVNGHKFSVSGEGDGTGKLTLCFTTGKLPVDPWPT 60

QY 61 LVTALSYGVCFSRYPDHMKQHDFFKSPMEGYVOERTIFFKDGNGYKTRAEVFEQDGL 120  
Db 61 LVTLSYGVCFSRYPDHMKQHDFFKSPMEGYVOERTIFFKDGNGYKTRAEVFEQDGL 120

QY 1:21 VNRLELKGIDFKEDGNILGHKLEYNYNHNVYINADKQKGIKNFKIRHNIEDGSVOLA 180  
Db 1:21 VNRLELKGIDFKEDGNILGHKLEYNYNHNVYINADKQKGIKNFKIRHNIEDGSVOLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLTSQSALSKDPNEKRDHWMLLGVPVTAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDGPVLLPDNHYLTSQSALSKDPNEKRDHWMLLGVPVTAAGITLGMDELYK 239

RESULT 2	
ABR40352	
ID	ABR40352 standard; protein; 363 AA.
XX	
AC	ABR40352;
XX	
DT	08-JUL-2003 (first entry)
XX	
DE	Human amino acid sequence SEQ ID NO: 6.
XX	
KW	Human; heterologous conjugate; intracellular protein.
XX	
OS	Homo sapiens.
OS	Aequoria victoria.
XX	
PN	WO2003029827-A2.
XX	
PD	10-APR-2003.
XX	
PF	01-OCT-2002; 2002WO-DK000651.
XX	
PR	01-OCT-2001; 2001DK-00001433.
PR	11-OCT-2001; 2001US-0328896P.
XX	
XX	
PA	(BIOL-) BIOLMAGE AS.
XX	
PI	Terry ER, Nielsen SJ;

XX	WPI; 2003-430211/40.
DR	N-PSDB; ACC72604.
XX	
PT	Novel cell for identifying modulators of protein interaction, contains a B
FT	first conjugate comprising anchor protein, second conjugate having type B
FT	interactor protein and third conjugate with detectable group.
XX	
XX	Disclosure; Page 112-113; 118pp; English.
PS	
CC	The invention relates to a novel cell, comprising three heterologous
CC	conjugates (HC), a first HC (HC1) comprising an anchor protein that
CC	specifically binds to an internal structure within the cell conjugated to
CC	an interactor protein (IP) of type A, a second HC (HC2) comprising IP of
CC	type B conjugated to a first protein of interest, and a third HC (HC3)
CC	comprising a second protein of interest conjugated to detectable group.
CC	The cell is useful for detecting if a compound disrupts or induces the
CC	interaction between two intracellular proteins. The cell is also useful
CC	for screening compounds that modulate the interaction between two
CC	intracellular proteins. The present sequence is used in the
CC	exemplification of the invention
XX	
SQ	Sequence 363 AA;
	Query Match            99.7%;     Score 1269;   DB 6;   Length 363;
	Best Local Similarity   99.6%;   Pred. No. 8.1e-124;
	Matches 238; Conservative   0; Mismatches   1; Indels   0; Gaps   0;
Qy	1 MVSKGEELFTGVVPILVELDGDVNGHKFVSFGEGDATYGKLTKEICTTTGKLFPVPMPT 60
Dd	1 MVSKGEELFTGVVPILVELDGDVNGHKFVSFGEGDATYGKLTKEICTTTGKLFPVPMPT 60
Qy	61 LVVTALSVCQCFSRYPPDHMKQHDFPKSAMPEGYVOERTIIPKDDGNKYTRAEVKFEGLT 120
Dd	61 LVVTLSYGVCQCFGRYPDPHMQHDFPKSAMPEGYVOERTIIPKDDGNKYTRAEVKFEGLT 120
Qy	121 VNRIELKGIDFKEDGNILGHLENYNNSHNVIYIMADKKNGIKVNFKIRHNIEDGSVQLA 180
Dd	121 VNRIELKGIDFKEDGNILGHLENYNNSHNVIYIMADKKNGIKVNFKIRHNIEDGSVQLA 180
Qy	181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMHMLLGFVTAAGITLGMDELYK 239
Dd	181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMHMLLGFVTAAGITLGMDELYK 239
RESULT 3	
AAG55781	
ID	AAG65781 standard; protein; 893 AA.
XX	AAG65781;
XX	
DT	07-JAN-2002 (first entry)
XX	
DE	Amino acid sequence of HSPDE4A1-E222G fusion protein.
XX	
KW	PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;
KW	autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;
KW	fusion protein.
OS	Homo sapiens.
OS	Aequorea victoria.
PN	WO200179526-A2.
XX	
PD	25-OCT-2001.
XX	
PF	11-APR-2001; 2001WO-DK000264.
XX	
PR	17-APR-2000; 2000DK-00000651.
PR	29-MAY-2000; 2000DK-00000849.
XX	(BIOI-) BIOIMAGE AS.
XX	

```
PI Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
PI Praestegaard M;
XX
XX WPI; 2001-611727/70.
DR N-PSDB; AAI66852.
XX
PT Determining if a compound is a dislocator of PDE4 for identifying
PT compounds for treating CNS and inflammatory disease comprises identifying
PT compounds which remove PDE4 spots.
XX
XX Example 1; Page 156-160; 160pp; English.
XX
XX The invention relates to determining, if a compound, is a dislocator of
XX PDE4. The method comprises testing if the compound removes PDE4 spots,
XX which may optionally be induced by a Rolipram-like reference compound,
XX and testing if it inhibits the catalytic activity of the PDE4, where the
XX compound is a dislocator of PDE4, if it removes PDE spots and if it does
XX not inhibit the catalytic activity of PDE4. The method is useful for
XX identifying compounds useful for the treatment of diseases of the central
XX nervous system such as, depression and for the treatment of inflammatory
XX disease such as joint inflammation, Crohn's disease, inflammatory bowel
XX disease, respiratory diseases, chronic obstructive pulmonary disease
XX (COPD), including asthma, chronic bronchitis, pulmonary emphysema,
XX endotoxemic shock, toxic shock syndrome, systemic lupus erythematosus,
XX psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
XX infection. The use of a reagent that can mimic or reverse the effect of
XX the compound with affinity for the catalytic site on intracellular
XX distribution of the PDE for the preparation of a medicament. The present
XX sequence represents the amino acid sequence of a HSPDE4A1-E222G fusion
XX protein
XX
XX Sequence 893 AA;
XX
XX Query Match 99.7%; Score 1269; DB 4; Length 893;
XX Best Local Similarity 99.6%; Pred. No. 3.1e-123;
XX Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MWSKGEEELFTGVVPILVELDGVNKGKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
DB 655 MWSKGEEELFTGVVPILVELDGVNKGKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 714
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDYL 120
DB 715 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDYL 774
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 775 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 834
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 835 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 893
RESULT 4
AAG65782
ID AAG65782 standard; protein; 1132 AA;
XX
XX AAG65782;
XX
XX 07-JAN-2002 (first entry)
XX
XX Amino acid sequence of HSPDE4A4-E222G fusion protein.
XX
XX PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;
XX autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;
XX fusion protein.
XX
XX Homo sapiens.
XX
XX Aequorea victoria.
XX
XX WO200179526-A2.
XX
XX
```

```
PD 25-OCT-2001.
XX
XX 11-APR-2001; 2001WO-DK000264.
XX
XX 17-APR-2000; 2000DK-00000651.
XX 29-MAY-2000; 2000DK-00000849.
XX
XX (BIOI-) BIOIMAGE AS.
XX
XX Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
XX Praestegaard M;
PI
XX WPI; 2001-611727/70.
DR N-PSDB; AAI66853.
XX
XX Determining if a compound is a dislocator of PDE4 for identifying
XX compounds for treating CNS and inflammatory disease comprises identifying
XX compounds which remove PDE4 spots.
XX
XX Example 1; Page 162-167; 160pp; English.
XX
XX The invention relates to determining, if a compound, is a dislocator of
XX PDE4. The method comprises testing if the compound removes PDE4 spots,
XX which may optionally be induced by a Rolipram-like reference compound,
XX and testing if it inhibits the catalytic activity of the PDE4, where the
XX compound is a dislocator of PDE4, if it removes PDE spots and if it does
XX not inhibit the catalytic activity of PDE4. The method is useful for
XX identifying compounds useful for the treatment of diseases of the central
XX nervous system such as, depression and for the treatment of inflammatory
XX disease such as joint inflammation, Crohn's disease, inflammatory bowel
XX disease, respiratory diseases, chronic obstructive pulmonary emphysema,
XX endotoxemic shock, toxic shock syndrome, systemic lupus erythematosus,
XX psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
XX infection. The use of a reagent that can mimic or reverse the effect of
XX the compound with affinity for the catalytic site on intracellular
XX distribution of the PDE for the preparation of a medicament. The present
XX sequence represents the amino acid sequence of a HSPDE4A4-E222G fusion
XX protein
XX
XX Sequence 1132 AA;
XX
XX Query Match 99.7%; Score 1269; DB 4; Length 1132;
XX Best Local Similarity 99.6%; Pred. No. 4.5e-123;
XX Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MWSKGEEELFTGVVPILVELDGVNKGKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
DB 894 MWSKGEEELFTGVVPILVELDGVNKGKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 953
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDYL 120
DB 954 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDYL 1013
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 1014 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 1073
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 1074 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 1132
RESULT 5
AAE17517
ID AAE17517 standard; protein; 239 AA.
XX
XX AAE17517;
XX
XX 22-APR-2002 (first entry)
XX
XX Enhanced F64L jellyfish green fluorescent protein mutant.
XX
XX
```

KW Jellyfish; green fluorescent protein; GFP; protein redistribution;  
 KW cellular function; genetic reporter; mutant; Stoke's shift; mutain.  
 XX  
 OS Aequorea victoria.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 65 /notes "wild type Phe substituted with Leu; This  
 FT corresponds to position 64 in the wild type protein"  
 XX  
 XX WO200198338-A2.  
 XX  
 XX 27-DEC-2001.  
 XX  
 XX 18-JUN-2001; 2001WO-EP006848.  
 XX  
 XX 19-JUN-2000; 2000DK-00000953.  
 PR 20-JUN-2000; 2000US-0212681P.  
 PR 10-MAY-2001; 2001DK-00000739.  
 PR 10-MAY-2001; 2001US-0290170P.  
 XX  
 XX (BIOI-) BIOIMAGE AS.  
 PA  
 XX Bjorn SP, Pagliaro L, Thastrup O;  
 PI  
 XX WPI; 2002-098224/13.  
 DR N-PSDB; AAD28162.  
 DR  
 XX Novel fluorescent protein in in vitro assay for measuring protein kinase  
 FT activity or dephosphorylation activity, or for measuring protein  
 PT redistribution, has a green fluorescent protein with F64L and E222G  
 PT mutation.  
 PT  
 XX Example 1; Page 35; 41pp; English.  
 PS  
 XX The invention relates to a fluorescent protein derived from green  
 CC fluorescent protein (GFP) or its analogue. The GFP containing mutations  
 CC at F64L and E222G has a bigger compared to other GFP's making it very  
 CC suitable for high throughput screening due to better resolution. The  
 CC fluorescent protein is useful in invitro assays for measuring protein  
 CC kinase activity or dephosphorylation activity, or for measuring protein  
 CC redistribution. The fluorescent protein is useful in studying cellular  
 CC functions in living cells; as protein tags in transgenic animals, living  
 CC and fixed cells; organelle tags, secretion marker and genetic reporter.  
 CC The fluorescent protein is also useful as a cell or organelle integrity  
 CC marker, a marker for changes in cell morphology, as transfection marker,  
 CC and as a marker to be used in combination with fluorescence activated  
 CC cell sorting (FACS). The novel proteins can also be used as reporters to  
 CC monitor live or dead biomass of organisms, such as fungi. The fluorescent  
 CC protein is also useful as markers in transcriptional and translational  
 CC fusions for performing transposon vector mutagenesis and as a reporter  
 CC for bacterial detection. Transposons encoding the fluorescent protein are  
 CC useful for screening promoters and for tagging plasmids and chromosomes.  
 CC The fluorescent protein engineered into the genome of a phage is useful  
 CC for designing diagnostic tool. The present sequence is enhanced F64L  
 CC Jellyfish green fluorescent protein (GFP) mutant  
 XX  
 SQ Sequence 239 AA;

Query Match 99.1%; Score 1261; DB 5; Length 239;  
 Best Local Similarity 99.2%; Pred. No. 3e-123;  
 Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MWSKGELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYKLTILFICTTKLPVWPWT 60  
 |||||  
 DB 1 MWSKGELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYKLTILFICTTKLPVWPWT 60  
 |||||  
 QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKRAEVKPGDFTL 120  
 |||||  
 DB 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKRAEVKPGDFTL 120  
 |||||  
 QY 121 VNRIELKGDIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDSGVQLA 180

DB 121 VNRIELKGDIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDSGVQLA 180  
 |||||  
 QY 181 DHYQQNTPIGDGPVLLPDNNHLSQTSALSQKDPNEKRDHVLGFVTAAGITLGMDELYK 239  
 |||||  
 DB 181 DHYQQNTPIGDGPVLLPDNNHLSQTSALSQKDPNEKRDHVLGFVTAAGITLGMDELYK 239  
 |||||

RESULT 6  
 AAB22882  
 ID AAB22882 standard; protein; 239 AA.  
 XX  
 AC AAB22882;  
 XX  
 DT 10-JAN-2001 (first entry)  
 XX  
 DE Enhanced green fluorescent protein (EGFP), SEQ ID NO:46.  
 XX  
 KW Bioreactor protein; fusion protein; recognition site;  
 KW cellular targeting sequence; cellular localisation; fluorescent protein;  
 KW protease activity detection; toxin detection; cellular stress detection;  
 KW drug discovery; cell based screening.  
 XX  
 OS Aequorea victoria.  
 OS Synthetic.  
 PN WO2000050872-A2.  
 XX  
 PD 31-AUG-2000.  
 XX  
 PR 25-FEB-2000; 2000WO-US004794.  
 XX  
 PR 26-FEB-1999; 99US-0122152P.  
 PR 08-MAR-1999; 99US-0023399P.  
 PR 12-JUL-1999; 99US-00352171.  
 XX  
 PA (CELL-) CELLOMICS INC.  
 XX  
 PI Giuliano KA, Kapur R;  
 XX  
 DR WPI; 2000-594086/56.  
 DR N-PSDB; AAA93373.  
 XX  
 PT Automated cell-based characterization of toxin by contacting cells  
 PT containing luminescent reporter molecules with test substance and  
 PT analyzing optically.  
 XX  
 PS Example 11; Fig 29A; 336pp; English.  
 CC  
 CC The invention relates to systems, methods and reagents for cell-based  
 CC screening or detection of compounds which affect particular biological  
 CC functions. The methods of the invention utilise fluorescent bioreactor  
 CC molecules which, when acted on by a compound of interest, cause an  
 CC alteration in the cellular distribution of at least the fluorescent  
 CC moiety. In one embodiment, the biosensors comprise heat shock proteins  
 CC (HSPs) fused to a fluorescent protein (e.g., jellyfish green fluorescent  
 CC protein (GFP), or derivatives thereof). Such biosensors are located in  
 CC the cytoplasm, but on stress activation translocate to the nucleus. In  
 CC another embodiment bioreactor proteins can be used to detect protease  
 CC activity. Such protease bioreactor fusion proteins comprise one or more  
 CC fluorescent proteins; a recognition signal which is cleaved by the  
 CC protease; and at least one cellular localisation signal. The latter two  
 CC components may be components of a single protein which is acted upon by  
 CC the protease, or may be from heterologous sources. Due to the  
 CC localisation signal, the bioreactor protein is localised to a particular  
 CC region of the cell. Once acted on by the protease of interest, the  
 CC fluorescent protein is cleaved from the localisation sequence, and is  
 CC free to migrate to other locations within the cell. The presence of a  
 CC second localisation signal attached to the fluorescent protein enables  
 CC the fluorescent protein to be directed to a different cellular  
 CC compartment after cleavage of the protease recognition sequence. The  
 CC change in distribution of the fluorescent protein can be detected using  
 CC imaging methods with a high degree of spatial resolution. The methods and



CC biosensors of the invention can be used to investigate a wide range of  
CC cellular activities and to screen compounds which modulate these  
CC activities. Biosensors containing a recognition site for caspase, for  
CC example, may be used for the screening of compounds which modulate  
CC apoptosis, while biosensors containing other protease recognition sites  
CC may be used for the detection of proteolytic toxins (such as anthrax  
CC lethal factor). The method provides improved target validation and  
CC candidate compound optimisation by combining many cell screening formats  
CC with fluorescence-based molecular reagents and computer-based feature  
CC extraction, data analysis and automation, resulting in increased quantity  
CC and speed of data collection and faster evaluation of drug candidates.  
CC Sequences AAB22881-B22885 represent fluorescent proteins which may be used  
CC as components of biosensor fusion proteins of the invention  
XX  
SQ Sequence 239 AA;

Query Match 98.8%; Score 1258; DB 3; Length 239;  
Best Local Similarity 98.7%; Pred. No. 6.1e-123;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 60  
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 7  
AAY54349  
ID AAY54349 standard; protein; 239 AA.  
XX AAY54349;  
AC AAY54349;  
XX  
DT 06-APR-2000 (first entry)  
XX  
DE Amino acid sequence of the mutant green fluorescent protein EGFP.  
XX  
XX Fluorescent protein; green fluorescent protein; emission intensity;  
KW fluorescence; pH detection; pH sensor; EGFP.  
XX  
XX Synthetic.  
OS Aequorea victoria.  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 65  
FT /note= "wild type Phe substituted with Leu"  
FT Misc-difference 66  
FT /note= "wild type Ser substituted with Thr"  
FT Misc-difference 232  
FT /note= "wild type His substituted with Leu"  
XX  
XX WO9964592-A2.  
PN  
XX  
PD 16-DEC-1999.  
XX  
XX 08-JUN-1999; 99WO-US012850.  
XX  
XX 09-JUN-1998; 98US-00094359.  
PR 13-OCT-1998; 98US-00172063.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
PA (UYOR-) UNIV OREGON STATE.  
XX

PI Tsien RY, Llopis J, Wachter RM;  
XX WPI; 2000-116540/10.  
DR N-PSDB; AA245642.  
XX  
PT New functional engineered green fluorescent proteins, used for measuring  
PT the pH in biological samples and cells.  
XX  
PS Disclosure; Page 9; 89pp; English.  
XX  
CC The present sequence represents a functional engineered fluorescent  
CC protein based on the Aequorea green fluorescent protein (GFP). The  
CC emission intensity changes as pH varies between 5 and 10 of the present  
CC protein are novel. The functional engineered fluorescent proteins show  
CC reversible changes in fluorescence over physiological pH ranges. They can  
CC be used for determining the pH of samples and cells. The polynucleotides  
CC can also be used to produce transgenic animals. The fluorescent protein  
CC pH sensors can be delivered to cells in the form of polynucleotides  
CC encoding the protein sensor fused to a targeting signal. The targeting  
CC signal directs the expression of the protein sensors to restricted cell  
CC locations. This makes it possible to measure the pH of a precisely  
CC defined cellular region or organelle  
XX  
SQ Sequence 239 AA;

Query Match 98.8%; Score 1258; DB 3; Length 239;  
Best Local Similarity 98.7%; Pred. No. 6.1e-123;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 60  
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 8  
AAY79584  
ID AAY79584 standard; peptide; 239 AA.  
XX  
AC AAY79584;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE EGFP signal domain.  
XX  
KW Protease; biosensor; EGFP; signal peptide; cell screening; assay;  
KW analysis; drug discovery.  
XX  
XX Unidentified.  
OS  
XX  
XX WO200026408-A2.  
PN  
XX  
PD 11-MAY-2000.  
XX  
XX 29-OCT-1999; 99WO-US025431.  
PF  
XX  
XX 30-OCT-1998; 98US-0106308P.  
PR 26-MAY-1999; 99US-0136078P.  
XX  
XX (CELL-) CELLOMICS INC.  
XX  
XX Giuliano KA, Bright G, Olson K, Burroughs-Tencza S;



```
PD 20-SEP-2001.
XX
PF 14-MAR-2001; 2001WO-US008071.
XX
PR 15-MAR-2000; 2000US-0189698P.
XX
XX (MERI ) MERCK & CO INC.
PA
XX Marsh DJ;
XX
XX WPI; 2001-565791/63.
XX
DR N-PSDB; AAH47304.
XX
XX Fusion proteins comprising melanin concentrating hormone receptor
PT peptides and fluorescent proteins, useful for identifying appetite
PT stimulants.
XX
PS Claim 2; Page 14; 71pp; English.
XX
CC The invention provides melanin concentrating hormone (MCH) receptor
CC (MCHR) chimeric and fusion proteins. The MCHR chimeric proteins comprise
CC MCHR polypeptide regions from different species. The MCHR fusion protein
CC comprise MCHR polypeptide region and a fluorescent polypeptide region
CC joined directly, or via a linker, to the carboxy side of the MCHR
CC polypeptide region. The MCHR fusion proteins can be expressed by standard
CC recombinant methodology. MCH action promotes feeding (orexigenic) and up
CC regulation of MCH activity stimulates food intake. The present sequence
CC represents a A. victoria green fluorescent protein (GFP) and a linker
CC sequence
XX
SQ Sequence 239 AA;

Query Match 98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 6.1e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTLKFICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTLKFICTTGKLPVWPWT 60

QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFDKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFDKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180

QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 11
AAB31171
ID AAB31171 standard; protein; 239 AA.
XX
AC AAB31171;
XX
DT 02-APR-2001 (first entry)
XX
DE Amino acid sequence of a green fluorescent protein (GFP).
XX
KW Growth rate; death rate; reporter gene; luminescent protein;
KW fluorescent product; luciferase; green fluorescent protein; GFP.
XX
OS Aequorea victoria.
XX
PN WO200075367-A1.
XX
PD 14-DEC-2000.
XX
PF 07-JUN-2000; 2000WO-FI000507.

XX 07-JUN-1999; 99FI-00001296.
XX
XX (LILI/) LILIUS E.
PA (VIRT/) VIRTA M.
XX
XX Lilius E, Virta M;
PI
XX WPI; 2001-061737/07.
DR N-PSDB; AAC86954.
XX
XX Assessing growth and death rates of a micro-organism in a desired
PT environment, by introducing 2 reporter genes encoding luminescent and
PT fluorescent products and detecting luminescent fluorescence.
XX
PS Disclosure; Page 27; 32pp; English.
XX
CC The specification describes a method for assessing the growth rate and
CC death rate of a micro-organism within a predetermined time period in a
CC desired environment. The method comprises introducing at least two
CC reporter genes encoding luminescent and/or fluorescent products into the
CC micro-organisms, incubating the micro-organism within the desired
CC environment, and detecting luminescence and/or fluorescence after a
CC predetermined time period. Use of two different markers within a micro-
CC organism enables the differentiation between growth and death rates. The
CC method is used to assess the growth rate and death rate of a micro-
CC organism within a predetermined time period in a desired environment. The
CC present sequence represents a green fluorescent protein (GFP), and is
CC encoded by a plasmid which encodes luminescent and fluorescent proteins,
CC and is used in the method of the invention
XX
SQ Sequence 239 AA;

Query Match 98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 6.1e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTLKFICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTLKFICTTGKLPVWPWT 60

QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFDKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFDKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180

QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 12
AAG66198
ID AAG66198 standard; protein; 239 AA.
XX
AC AAG66198;
XX
DT 17-JUN-2002 (first entry)
XX
DE A. victoria green fluorescent protein (EGFP).
XX
KW Cyan-green fluorescent protein; fluorescence; recombinant; GFP;
KW green fluorescent protein; EGFP.
XX
OS Aequorea victoria.
XX
PN JP2002045189-A.
XX
PD 12-FEB-2002.
XX
```

```

PF 04-AUG-2000; 2000JP-00237165.
XX
PR 04-AUG-2000; 2000JP-00237165.
XX
XX (RIKA ) RIKAGAKU KENKYUSHO.
XX
XX WPI; 2002-299190/34.
XX
XX N-PSDB; ABL40628.
XX
XX A gene encoding cyan-green fluorescent protein.
XX
XX
XX Example; Page 14; 20pp; Japanese.
XX
XX The invention relates to a gene encoding proteins having cyan-green
XX fluorescence characteristic and having a function of showing stable
XX fluorescence characteristic in acid region. A method for the preparation
XX of a cyan-green fluorescent protein is provided which involves a
XX transforment transformed by a recombinant vector comprising the gene,
XX where the transformant is cultured and the protein is collected from the
XX culture. The present sequence represents the A. victoria green
XX fluorescent protein (EGFP)
XX
XX Sequence 239 AA;
SQ
Query Match 98.8%; Score 1258; DB 5; Length 239;
Best Local Similarity 98.7%; Pred. No. 6.1e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MWSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTCLKFICTTGKLPVWPWT 60
Db 1 MWSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTCLKFICTTGKLPVWPWT 60
QY 51 LVTALSYGQCFSRYPDHMKQHDFFKPSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 51 LVTTLTYGVQCFSRYPDHMKQHDFFKPSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 131 DHYQNTPTIGDGPVLLPDNHYLSTQSALSCKPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
Db 131 DHYQNTPTIGDGPVLLPDNHYLSTQSALSCKPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
RESULT 13
ABG94444
ID ABG94444 standard; protein; 239 AA.
XX
XX ABG94444;
XX
XX 27-NOV-2002 (first entry)
XX
XX Protease biosensor signal sequence #6.
XX
XX Detection; classification; identification; toxin detection; protease;
XX ADP-ribosylating toxin; cytotoxic phospholipase; exfoliative toxin;
XX toxic threat agent.
XX
XX Synthetic.
XX
XX US6416959-B1.
XX
XX 09-JUL-2002.
XX
XX 25-FEB-2000; 2000US-00513783.
XX
XX 27-FEB-1997; 97US-00810983.
XX
XX 27-FEB-1998; 98US-00031271.
XX
XX 26-FEB-1999; 99US-0122152P.
XX
XX 08-MAR-1999; 99US-0123399P.
XX
XX 12-JUL-1999; 99US-00352171.
XX
XX 31-AUG-1999; 99US-0151797P.
XX
PR 17-SEP-1999; 99US-00398965.
PR 29-OCT-1999; 99US-00430656.
PR 01-DEC-1999; 99US-0168408P.
XX
XX (GIUL/) GIULIANO K.
XX (KAPU/) KAPUR R.
XX
XX Giuliano K, Kapur R;
XX
XX WPI; 2002-634730/68.
XX
XX N-PSDB; ABS71491.
XX
XX Automated cell-based toxin detection, classification, and/or
XX identification by treating cells involves use of three classes of
XX luminescent reporter molecules such as detectors, classifiers or
XX identifiers.
XX
XX Example 10; Fig 29A; 214pp; English.
XX
XX The invention describes methods of automated detection, classification
XX and identification comprising treating cells containing luminescent
XX reporter molecules (1) in array of locations with a test substance, where
XX (1) are detectors, classifiers or identifiers, imaging cells in each
XX location to obtain luminescent signals and converting optical information
XX into digital data to interpret presence of toxins in the test substance.
XX The method are useful for detection of toxins chosen from proteases, ADP-
XX ribosylating toxins, cytotoxic phospholipases, and exfoliative toxins.
XX Three classes of cell-based luminescent reporter molecules such as
XX detectors, classifiers and identifiers are described and serve as
XX reporters of toxic threat agents. The first two levels of
XX characterisation ensure a rapid readout of toxin class without
XX sacrificing the ability to detect many new mutant toxins or dissect
XX several complex mixtures of known toxins. This is the amino acid sequence
XX of a protease biosensor related signal sequence used in the cell-based
XX screening system
XX
XX Sequence 239 AA;
SQ
Query Match 98.8%; Score 1258; DB 5; Length 239;
Best Local Similarity 98.7%; Pred. No. 6.1e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MWSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTCLKFICTTGKLPVWPWT 60
Db 1 MWSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTCLKFICTTGKLPVWPWT 60
QY 61 LVTALSYGQCFSRYPDHMKQHDFFKPSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLTYGVQCFSRYPDHMKQHDFFKPSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSCKPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSCKPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
RESULT 14
AAE14599
ID AAE14599 standard; protein; 239 AA.
XX
XX AAE14599;
XX
XX 31-MAY-2002 (first entry)
XX
XX Aequorea victoria enhanced green fluorescent protein.
XX
XX Mutagenesis; enhanced green fluorescent protein; EGFP; mutant; mutein.
XX
XX Aequorea victoria.
XX
XX Synthetic.

```

XX Key Location/Qualifiers  
 FT Misc-difference 1. .3 /note= "Wild-type GFP Met-Ser are replaced with Met-Val-Ser"  
 FT Misc-difference 65 /note= "GFP Phe64 is replaced by Leu"  
 FT Misc-difference 66 /note= "GFP Ser65 is replaced by Thr"  
 XX EP1178109-A1.  
 XX 06-FEB-2002.  
 XX 03-AUG-2001; 2001EP-00306650.  
 XX 04-AUG-2000; 2000JP-00237166.  
 XX (RIKE ) RIKEN KK.  
 XX Miyawaki A, Sawano A;  
 XX WPI; 2002-208112/27.  
 XX N-PSDB; AAD27910.  
 XX Method for mutagenesis, e.g. for introducing certain or random mutations at certain sites of the nucleotide sequence, comprises synthesizing a mutated strand and a complementary strand by use of megaprimers.  
 XX Example 1; Page 13-14; 31pp; English.  
 XX The invention relates to a method for mutagenesis that comprises synthesizing a mutated strand and a complementary strand by use of megaprimers. The method basically comprises a DNA synthesis in which one or more primers that have a nucleotide sequence containing at least one mutation and a phosphorylated 5'-terminus are annealed to a template DNA and then subjected to an elongation reaction using a thermostable high-fidelity DNA polymerase, after which the phosphorylated 5'-terminus and the elongated terminus are ligated by means of a thermostable DNA ligase to synthesise a circular DNA containing the primers; a digestion in which the step of DNA synthesis is repeated several times to amplify the DNA containing the primers and then, at least DNAs other than the amplified circular DNA are digested into several fragments; and a double-stranded DNA synthesis in which, with the several fragments obtained in the step of digestion as megaprimers, the megaprimers are annealed to the circular DNA synthesised above, followed by an elongation reaction performed using the thermostable high-fidelity DNA polymerase. The method is useful for mutagenesis, particularly for introducing certain mutations at certain sites of a nucleotide sequence, or for introducing random mutations at certain sites of the nucleotide sequence. The present method is simple, speedy, economical and widely applicable. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used for mutagenesis in an exemplification of the invention. The EGFP is derived by altering the green fluorescent protein (GFP) sequence of Aequorea victoria  
 XX Sequence 239 AA;  
 Query Match 98.8%; Score 1258; DB 5; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 6.1e-123;  
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60  
 DB 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60  
 QY 61 LVTALSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTLLTYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGDIFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 DB 121 VNRIELKGDIFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 QY 181 DHYQQNTPTIGDGVLLPDNHYLSTQSALSKDPNEKRDHMLVGLFVTAAGITLGMDELYK 239  
 DB 181 DHYQQNTPTIGDGVLLPDNHYLSTQSALSKDPNEKRDHMLVGLFVTAAGITLGMDELYK 239

QY 181 DHYQQNTPTIGDGVLLPDNHYLSTQSALSKDPNEKRDHMLVGLFVTAAGITLGMDELYK 239  
 DB 181 DHYQQNTPTIGDGVLLPDNHYLSTQSALSKDPNEKRDHMLVGLFVTAAGITLGMDELYK 239  
 RESULT 15  
 AAE34958  
 ID AAE34958 standard; protein; 239 AA.  
 XX AAE34958;  
 AC AAE34958;  
 XX 28-MAY-2003 (first entry)  
 XX Aequorea victoria enhanced green fluorescent protein (EGFP).  
 XX Phosphorylation indicator; fluorescent protein; detection; phosphatase;  
 KW Kinase; enhanced green fluorescent protein; EGFP.  
 OS Aequorea victoria.  
 XX WO200295058-A2.  
 XX 28-NOV-2002.  
 XX 24-MAY-2002; 2002WO-US016955.  
 XX 24-MAY-2001; 2001US-00865291.  
 XX (REGC ) UNIV CALIFORNIA.  
 XX Tsien RY, Ting AY, Zhang J;  
 WPI; 2003-148474/14.  
 N-PSDB; AAD53428.  
 XX Novel chimeric phosphorylation indicators, useful for detecting kinase/phosphatase in samples, has donor molecule, phosphorylatable domain, phosphaminoacid binding domain, and acceptor molecule, in operative linkage.  
 PS Disclosure; Col 56-57; 38pp; English.  
 XX The present invention relates to chimeric phosphorylation indicators comprising a phosphorylation polypeptide and a fluorescent protein or operative linkage, a donor molecule, a phosphorylatable domain, a phosphaminoacid binding domain (PAABD) and an acceptor molecule. The phosphorylation indicators of the invention are useful for detecting kinases or phosphatases in a biological sample. They are also useful in high throughput analysis e.g. for detecting a kinase inhibitor or phosphatase inhibitor. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used in the invention  
 XX Sequence 239 AA;  
 Query Match 98.8%; Score 1258; DB 6; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 6.1e-123;  
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60  
 DB 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60  
 QY 61 LVTALSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTLLTYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGDIFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 DB 121 VNRIELKGDIFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 QY 181 DHYQQNTPTIGDGVLLPDNHYLSTQSALSKDPNEKRDHMLVGLFVTAAGITLGMDELYK 239  
 DB 181 DHYQQNTPTIGDGVLLPDNHYLSTQSALSKDPNEKRDHMLVGLFVTAAGITLGMDELYK 239

us-09-887-784-64a.rag

Sun Jun 27 18:27:35 2004

Db 181 DHYQONTPIGDGEVLLPDNHYLSTQSALSKDPNEKRDHVVLEFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 15:54:09  
Job time : 48.1111 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:55 ; Search time 12.7778 Seconds  
(without alignments)  
965.630 Million cell updates/sec

Title: US-09-887-784-64A

Perfect score: 1273  
Sequence: 1 MWSKGEELFTGVVPIVLVD.....VLLGFVTRAGITGMDELYK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/prodata/2/iaa/5A COMB.pep.\*
- 2: /cgn2\_6/prodata/2/iaa/5B COMB.pep.\*
- 3: /cgn2\_6/prodata/2/iaa/6A COMB.pep.\*
- 4: /cgn2\_6/prodata/2/iaa/6B COMB.pep.\*
- 5: /cgn2\_6/prodata/2/iaa/PCTUS COMB.pep.\*
- 6: /cgn2\_6/prodata/2/iaa/backfilese1.pep.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1258	98.8	239	3	US-09-172-063-3
2	1258	98.8	239	4	US-09-513-783A-46
3	1258	98.8	239	4	US-09-316-919-4
4	1258	98.8	239	4	US-09-602-641-3
5	1258	98.8	239	4	US-09-920-922-2
6	1258	98.8	281	3	US-09-062-102-1
7	1258	98.8	281	4	US-09-364-946-1
8	1258	98.8	294	4	US-09-513-783A-2
9	1258	98.8	323	3	US-09-172-063-21
10	1258	98.8	323	4	US-09-602-641-21
11	1258	98.8	364	3	US-09-085-305-6
12	1258	98.8	379	4	US-09-417-197-129
13	1258	98.8	434	4	US-09-800-170-48
14	1258	98.8	442	4	US-09-417-197-127
15	1258	98.8	459	4	US-09-513-783A-170
16	1258	98.8	544	4	US-09-417-197-113
17	1258	98.8	544	4	US-09-417-197-115
18	1258	98.8	604	4	US-09-417-197-59
19	1258	98.8	605	4	US-09-417-197-41
20	1258	98.8	606	4	US-09-417-197-65
21	1258	98.8	607	4	US-09-417-197-47
22	1258	98.8	630	4	US-09-417-197-63
23	1258	98.8	631	4	US-09-417-197-39
24	1258	98.8	633	4	US-09-417-197-45
25	1258	98.8	635	4	US-09-417-197-125
26	1258	98.8	642	2	US-08-818-253-2
27	1258	98.8	642	2	US-08-818-253-6

RESULT 1  
US-09-172-063-3  
; Sequence 3, Application US/09172063  
; Patent No. 6150176  
; GENERAL INFORMATION:  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Miyawaki, Atsushi  
; APPLICANT: Liopis, Juan  
; APPLICANT: Wachter, Rebekka M.  
; APPLICANT: Remington, S. James  
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
; FILE REFERENCE: 07257/071001  
; CURRENT APPLICATION NUMBER: US/09/172,063  
; CURRENT FILING DATE: 1998-10-13  
; EARLIER APPLICATION NUMBER: 09/094,359  
; EARLIER FILING DATE: 1998-06-09  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: EGFP  
US-09-172-063-3

Query Match 98.8%; Score 1258; DB 3; Length 239;  
Best Local Similarity 98.7%; Pred. No. 2.7e-127;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MWSKGEELFTGVVPIVLVDGVNKHFSVSGEGDATYVKLTLPKFTCTTGKLPVPWPT 60  
DB 1 MWSKGEELFTGVVPIVLVDGVNKHFSVSGEGDATYVKLTLPKFTCTTGKLPVPWPT 60  
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVYQVETIFFKDDGNKTRAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQVETIFFKDDGNKTRAEVKFEGDTL 120  
QY 121 VNRIELKIDFKEGDNILGHKLEYNVNSHNVIYIMADKQNGIKVNFIRHNIEDGSVOLA 180  
DB 121 VNRIELKIDFKEGDNILGHKLEYNVNSHNVIYIMADKQNGIKVNFIRHNIEDGSVOLA 180  
QY 181 DHYQONTPTGDCPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPTGDCPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

ALIGNMENTS

28 1258 98.8 642 3 US-08-818-252-2 Sequence 2, Appli  
29 1258 98.8 642 3 US-08-818-252-6 Sequence 6, Appli  
30 1258 98.8 652 2 US-08-818-253-4 Sequence 4, Appli  
31 1258 98.8 652 3 US-08-818-253-4 Sequence 4, Appli  
32 1258 98.8 718 4 US-09-417-197-75 Sequence 75, Appli  
33 1258 98.8 719 4 US-09-417-197-51 Sequence 51, Appli  
34 1258 98.8 726 4 US-09-417-197-71 Sequence 71, Appli  
35 1258 98.8 727 4 US-09-417-197-139 Sequence 139, App  
36 1258 98.8 783 4 US-09-513-783A-176 Sequence 176, App  
37 1258 98.8 797 4 US-09-417-197-141 Sequence 141, App  
38 1258 98.8 797 4 US-09-417-197-143 Sequence 143, App  
39 1258 98.8 798 4 US-09-417-197-77 Sequence 77, Appli  
40 1258 98.8 805 4 US-09-513-783A-178 Sequence 178, App  
41 1258 98.8 806 4 US-09-417-197-53 Sequence 53, Appli  
42 1258 98.8 836 4 US-09-417-197-61 Sequence 61, Appli  
43 1258 98.8 842 4 US-09-417-197-43 Sequence 43, Appli  
44 1258 98.8 843 4 US-09-417-197-117 Sequence 117, App  
45 1258 98.8 853 4 US-09-417-197-119 Sequence 119, App

```
RESULT 2
US-09-513-733A-46
; Sequence 46, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGFP
US-09-513-733A-46

Query Match          98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.7e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYKLTLLKFKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYKLTLLKFKICTTGKLPVPWPT 60
QY 51 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 51 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 131 DHYQNTPTIGDGPVLLPDNHYLSTQALSCKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 131 DHYQNTPTIGDGPVLLPDNHYLSTQALSCKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 3
US-09-316-919-4
; Sequence 4, Application US/09316919
; Patent No. 6469154
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: 07257/073001
; CURRENT APPLICATION NUMBER: US/09/316,919
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-316-919-4

Query Match          98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.7e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYKLTLLKFKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYKLTLLKFKICTTGKLPVPWPT 60
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 131 DHYQNTPTIGDGPVLLPDNHYLSTQALSCKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 131 DHYQNTPTIGDGPVLLPDNHYLSTQALSCKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 4
US-09-602-641-3
; Sequence 3, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Ilopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-602-641-3

Query Match          98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.7e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYKLTLLKFKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYKLTLLKFKICTTGKLPVPWPT 60
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 131 DHYQNTPTIGDGPVLLPDNHYLSTQALSCKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 131 DHYQNTPTIGDGPVLLPDNHYLSTQALSCKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 5
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. 6673610
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
```



```

; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match      98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.7e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60
   |||||
DB 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60
   |||||

QY 61 LVTALSYGVCFSRYPDHMKQHDFFKFSAMPEGVYQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
   |||||
DB 61 LVTLTLYGVCFSRYPDHMKQHDFFKFSAMPEGVYQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
   |||||

QY 121 VNRIELKGIDFKEDGNILGHKLEYNYSNNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
   |||||
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYSNNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
   |||||

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
   |||||
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
   |||||

RESULT 6
US-09-062-102-1
; Sequence 1, Application US/09062102
; Patent No. 6130313
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangqiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100
; CURRENT APPLICATION NUMBER: US/09/062,102
; CURRENT FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: US 60/060,855
; EARLIER FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6130313
US-09-062-102-1

Query Match      98.8%; Score 1258; DB 3; Length 281;
Best Local Similarity 98.7%; Pred. No. 3.4e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60
   |||||
DB 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60
   |||||

QY 61 LVTALSYGVCFSRYPDHMKQHDFFKFSAMPEGVYQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
   |||||
DB 61 LVTLTLYGVCFSRYPDHMKQHDFFKFSAMPEGVYQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
   |||||

QY 121 VNRIELKGIDFKEDGNILGHKLEYNYSNNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
   |||||
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYSNNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
   |||||

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
   |||||
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
   |||||

RESULT 7
US-09-364-946-1
; Sequence 1, Application US/09364946
; Patent No. 6306600
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangqiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100CIP/D2
; CURRENT APPLICATION NUMBER: US/09/364,946
; CURRENT FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: US 09/191,233
; EARLIER FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6306600
US-09-364-946-1

Query Match      98.8%; Score 1258; DB 4; Length 281;
Best Local Similarity 98.7%; Pred. No. 3.4e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60
   |||||
DB 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60
   |||||

QY 61 LVTALSYGVCFSRYPDHMKQHDFFKFSAMPEGVYQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
   |||||
DB 61 LVTLTLYGVCFSRYPDHMKQHDFFKFSAMPEGVYQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
   |||||

QY 121 VNRIELKGIDFKEDGNILGHKLEYNYSNNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
   |||||
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYSNNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
   |||||

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
   |||||
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
   |||||

RESULT 8
US-09-513-783A-2
; Sequence 2, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GFP-DEVD-Annexin II construct
US-09-513-783A-2

Query Match      98.8%; Score 1258; DB 4; Length 294;
Best Local Similarity 98.7%; Pred. No. 3.7e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60
   |||||
DB 1 MYSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60
   |||||
```

Db 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYVKLTLEFKICTTGKLPVWPWT 60  
QY 51 LVTALSYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 51 LVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
QY 191 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239  
Db 191 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239

## RESULT 9

US-09-172-053-21

; Sequence 21, Application US/09172063

; Patent No. 6150176

; GENERAL INFORMATION:

; APPLICANT: Tsien, Roger Y.

; APPLICANT: Miyawaki, Atsushi

; APPLICANT: Llopis, Juan

; APPLICANT: Wachter, Rebekka M.

; APPLICANT: Remington, S. James

; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR

; FILE REFERENCE: 07257/071001

; CURRENT APPLICATION NUMBER: US/09/172,063

; CURRENT FILING DATE: 1998-10-13

; EARLIER APPLICATION NUMBER: 09/094,359

; EARLIER FILING DATE: 1998-06-09

; NUMBER OF SEQ ID NOS: 38

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 21

; LENGTH: 323

; TYPE: PRT

; ORGANISM: Aequorea victoria

; FEATURE:

; NAME/KEY: VARIANT

; LOCATION: (0)...(0)

; OTHER INFORMATION: GT-EGFP

US-09-172-053-21

Query Match 98.8%; Score 1258; DB 3; Length 323;

Best Local Similarity 98.7%; Pred. No. 4.2e-127;

Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYVKLTLEFKICTTGKLPVWPWT 60  
Db 35 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYVKLTLEFKICTTGKLPVWPWT 144  
QY 51 LVTALSYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 145 LVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 204  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
Db 205 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 264  
QY 191 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239  
Db 255 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 323

## RESULT 10

US-09-602-641-21

; Sequence 21, Application US/09602641

; Patent No. 6608189

; GENERAL INFORMATION:

; APPLICANT: Tsien, Roger Y.

; APPLICANT: Miyawaki, Atsushi

; APPLICANT: Llopis, Juan

; APPLICANT: Wachter, Rebekka M.  
; APPLICANT: Remington, S. James  
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
; FILE REFERENCE: 07257/071001  
; CURRENT APPLICATION NUMBER: US/09/602,641  
; CURRENT FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 09/172,063  
; PRIOR FILING DATE: 1998-10-13  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 21  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: GT-EGFP  
US-09-602-641-21

Query Match 98.8%; Score 1258; DB 4; Length 323;

Best Local Similarity 98.7%; Pred. No. 4.2e-127;

Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYVKLTLEFKICTTGKLPVWPWT 60  
Db 85 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYVKLTLEFKICTTGKLPVWPWT 144  
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 145 LVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 204  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
Db 205 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 264  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239  
Db 265 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 323

## RESULT 11

US-09-085-305-6

; Sequence 6, Application US/09085305

; Patent No. 6191269

; GENERAL INFORMATION:

; APPLICANT: Pollock, Allan

; APPLICANT: Lovett, David H.

; APPLICANT: Turck, Johanna

; TITLE OF INVENTION: Selective Induction of Apoptosis in

; TITLE OF INVENTION: Malignant Cancer Cells by Delivery of N-Terminal

; TITLE OF INVENTION: Interleukin-1-Alpha Pro-Piece Polypeptide

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Bozicevic &amp; Reed, LLP

; STREET: 285 Hamilton Ave, Suite 200

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94301

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/085,305

; FILING DATE: 29-MAY-1998

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Francis, Carol L  
REGISTRATION NUMBER: 36,513  
REFERENCE/DOCKET NUMBER: 6510/102US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-327-3400  
TELEFAX: 650-327-3231  
TELEX:  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 364 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-085-305-6

Query Match 98.8%; Score 1258; DB 3; Length 364;  
Best Local Similarity 98.7%; Pred. No. 5.1e-127;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPKICTTGKLPVPWPT 60  
DB 126 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPKICTTGKLPVPWPT 185  
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVYQVETIFFKDDGNYKTRAEVKFEGDTL 120  
DB 186 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQVETIFFKDDGNYKTRAEVKFEGDTL 245  
QY 121 VNRLEKIGDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 246 VNRLEKIGDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 305  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB 306 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 364

RESULT 12  
US-09-417-197-129  
Sequence 129, Application US/09417197  
Patent No. 6518021  
GENERAL INFORMATION:  
APPLICANT: Ole THASTRUP, et al.  
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An  
FILE REFERENCE: 3759-0110P  
CURRENT FILING DATE: 1999-10-07  
NUMBER OF SEQ ID NOS: 143  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 129  
LENGTH: 379  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: actin-binding-domain-EGFP fusion  
US-09-417-197-129

Query Match 98.8%; Score 1258; DB 4; Length 379;  
Best Local Similarity 98.7%; Pred. No. 5.4e-127;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPKICTTGKLPVPWPT 60  
DB 141 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPKICTTGKLPVPWPT 200  
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVYQVETIFFKDDGNYKTRAEVKFEGDTL 120  
DB 201 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQVETIFFKDDGNYKTRAEVKFEGDTL 260  
QY 121 VNRLEKIGDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 261 VNRLEKIGDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 320

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB 321 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 379

RESULT 13  
US-09-800-170-48  
Sequence 48, Application US/09800170  
Patent No. 6481667  
GENERAL INFORMATION:  
APPLICANT: Kinsella, Todd  
TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES  
FILE REFERENCE: A-68614-1/DJB/RMS/RMK  
CURRENT APPLICATION NUMBER: US/09/800,170  
CURRENT FILING DATE: 2001-08-28  
PRIOR APPLICATION NUMBER: US 60/187,130  
PRIOR FILING DATE: 2000-03-06  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 48  
LENGTH: 434  
TYPE: PRT  
ORGANISM: Synechocystis PCC6803  
US-09-800-170-48

Query Match 98.8%; Score 1258; DB 4; Length 434;  
Best Local Similarity 98.7%; Pred. No. 6.6e-127;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPKICTTGKLPVPWPT 60  
DB 196 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPKICTTGKLPVPWPT 255  
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVYQVETIFFKDDGNYKTRAEVKFEGDTL 120  
DB 256 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQVETIFFKDDGNYKTRAEVKFEGDTL 315  
QY 121 VNRLEKIGDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 316 VNRLEKIGDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 375  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB 376 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 434

RESULT 14  
US-09-417-197-127  
Sequence 127, Application US/09417197  
Patent No. 6518021  
GENERAL INFORMATION:  
APPLICANT: Ole THASTRUP, et al.  
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An  
FILE REFERENCE: 3759-0110P  
CURRENT FILING DATE: 1999-10-07  
NUMBER OF SEQ ID NOS: 143  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 127  
LENGTH: 442  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: EGFP-RhoA fusion  
US-09-417-197-127

Query Match 98.8%; Score 1258; DB 4; Length 442;  
Best Local Similarity 98.7%; Pred. No. 6.8e-127;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPKICTTGKLPVPWPT 60

Db 1 MVSKGEELFTGVVPIVLVELDGVNGHKFSVSGEGGDATYGKLTLCFICTTGKLPVPWPT 60  
Qy 61 LVTALSYGVQCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKPEGDTL 120  
Db 61 LVTTLTYGVQCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKPEGDTL 120  
Qy 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180  
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180  
Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 15  
US-09-513-783A-170  
; Sequence 170, Application US/09513783A  
; Patent No. 6416959  
; GENERAL INFORMATION:  
; APPLICANT: Giuliano, Kenneth A.  
; APPLICANT: Kapur, Ravi  
; TITLE OF INVENTION: A System for Cell Based Screening  
; FILE REFERENCE: 97-022-11  
; CURRENT APPLICATION NUMBER: US/09/513,783A  
; CURRENT FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 170  
; LENGTH: 459  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP27  
US-09-513-783A-170

Query Match 98.8%; Score 1258; DB 4; Length 459;  
Best Local Similarity 98.7%; Pred. No. 7.2e-127; Indels 0; Gaps 0;  
Matches 236; Conservative 1; Mismatches 2;  
Qy 1 MVSKGEELFTGVVPIVLVELDGVNGHKFSVSGEGGDATYGKLTLCFICTTGKLPVPWPT 60  
Db 1 MVSKGEELFTGVVPIVLVELDGVNGHKFSVSGEGGDATYGKLTLCFICTTGKLPVPWPT 60  
Qy 61 LVTALSYGVQCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKPEGDTL 120  
Db 61 LVTTLTYGVQCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKPEGDTL 120  
Qy 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180  
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180  
Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:04:01  
Job time : 13.778 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:54:20 ; Search time 34.7778 Seconds  
(without alignments)  
1940.117 Million cell updates/sec

Title: US-09-887-784-64A

Perfect score: 1273

Sequence: 1 MVSKEELFTGVVPILVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US05\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1269	99.7	239	9	US-09-887-784-4
2	1269	99.7	239	12	US-10-296-953-4
3	1269	99.7	363	14	US-10-270-223-6
4	1269	99.7	893	14	US-10-257-909A-30
5	1269	99.7	1132	14	US-10-257-909A-32
6	1261	99.1	239	9	US-09-887-784-2
7	1261	99.1	239	12	US-10-296-953-2
8	1258	98.8	239	9	US-09-920-922-2
9	1258	98.8	239	9	US-09-959-745-4
10	1258	98.8	239	10	US-09-866-538-4
11	1258	98.8	239	10	US-09-797-496B-2
12	1258	98.8	239	10	US-09-794-308-4
13	1258	98.8	239	10	US-09-865-291-4
14	1258	98.8	239	12	US-10-457-982-3
15	1258	98.8	239	14	US-10-121-258-13

16	1258	98.8	239	14	US-10-221-461-7	Sequence 7, Appli
17	1258	98.8	239	14	US-10-100-957A-46	Sequence 46, Appli
18	1258	98.8	239	14	US-10-177-390-2	Sequence 2, Appli
19	1258	98.8	239	14	US-10-338-411-3	Sequence 3, Appli
20	1258	98.8	239	15	US-10-370-570-4	Sequence 4, Appli
21	1258	98.8	239	15	US-10-389-640-3	Sequence 3, Appli
22	1258	98.8	259	14	US-10-314-861-11	Sequence 11, Appli
23	1258	98.8	281	12	US-09-931-232-1	Sequence 11, Appli
24	1258	98.8	288	14	US-10-314-861-37	Sequence 37, Appli
25	1258	98.8	293	14	US-10-314-861-35	Sequence 35, Appli
26	1258	98.8	294	14	US-10-100-957A-2	Sequence 2, Appli
27	1258	98.8	295	14	US-10-314-861-39	Sequence 39, Appli
28	1258	98.8	299	14	US-10-314-861-33	Sequence 33, Appli
29	1258	98.8	305	14	US-10-314-861-31	Sequence 31, Appli
30	1258	98.8	308	14	US-10-033-717-35	Sequence 35, Appli
31	1258	98.8	311	14	US-10-314-861-29	Sequence 29, Appli
32	1258	98.8	320	14	US-10-338-411-11	Sequence 11, Appli
33	1258	98.8	320	15	US-10-389-640-11	Sequence 11, Appli
34	1258	98.8	323	12	US-10-457-982-21	Sequence 21, Appli
35	1258	98.8	323	14	US-10-338-411-7	Sequence 7, Appli
36	1258	98.8	323	14	US-10-338-411-13	Sequence 13, Appli
37	1258	98.8	323	15	US-10-389-640-7	Sequence 7, Appli
38	1258	98.8	323	15	US-10-389-640-13	Sequence 13, Appli
39	1258	98.8	324	14	US-10-314-861-16	Sequence 16, Appli
40	1258	98.8	345	14	US-10-338-411-5	Sequence 5, Appli
41	1258	98.8	345	15	US-10-389-640-5	Sequence 5, Appli
42	1258	98.8	346	14	US-10-338-411-9	Sequence 9, Appli
43	1258	98.8	346	15	US-10-389-640-9	Sequence 9, Appli
44	1258	98.8	359	14	US-10-033-717-33	Sequence 33, Appli
45	1258	98.8	359	14	US-10-033-717-34	Sequence 34, Appli

ALIGNMENTS

RESULT 1

US-09-887-784-4  
; Sequence 4, Application US/09887784  
; Patent No. US20020177189A1  
; GENERAL INFORMATION:  
; APPLICANT: BJORN, Sara et al  
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS  
; FILE REFERENCE: 3759-0115P  
; CURRENT APPLICATION NUMBER: US/09/887,784  
; CURRENT FILING DATE: 2001-06-19  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequoria Victoria  
US-09-887-784-4

Query Match 99.7%; Score 1269; DB 9; Length 239;  
Best Local Similarity 99.6%; Pred. No. 5.8e-124;  
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MVSKGBELFTGVVPILVELDGDVNGHKFSVSGEGGDATY	GKLT	KFI	CTTGKLPVPWPT	60
Db	1	MVSKGBELFTGVVPILVELDGDVNGHKFSVSGEGGDATY	GKLT	KFI	CTTGKLPVPWPT	60
Qy	61	LVTALSYGVQCFSRYPDHMKQDFFPKSAMPEGVYQERTIFF	KDDGNYKTRAEVKFEGD	TL	120	
Db	61	LVTTLSSYGVQCFSRYPDHMKQDFFPKSAMPEGVYQERTIFF	KDDGNYKTRAEVKFEGD	TL	120	
Qy	121	VNRIELKGIIDFKEDGNILGHLEYNVYNSHNVIMADKQKNGIK	VNFKIRHNIEDGSVQLA	180		
Db	121	VNRIELKGIIDFKEDGNILGHLEYNVYNSHNVIMADKQKNGIK	VNFKIRHNIEDGSVQLA	180		
Qy	181	DHYQQNTPTIGDGVLLPDNHYLSTQSALSKDPNEKRDHMLLGF	VTAAGITLGMDELYK	239		
Db	181	DHYQQNTPTIGDGVLLPDNHYLSTQSALSKDPNEKRDHMLLGF	VTAAGITLGMDELYK	239		

RESULT 2

US-10-296-953-4

Sequence 4, Application US/10296953

Publication No. US20040072995A1

GENERAL INFORMATION:

APPLICANT: BJORN, SARA P.

APPLICANT: PAGLIARO, LEN

APPLICANT: THASTRUP, OLE

TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS

FILE REFERENCE: PL0095

CURRENT APPLICATION NUMBER: US/10/296,953

CURRENT FILING DATE: 2002-11-26

PRIOR APPLICATION NUMBER: PA 2000 00953

PRIOR FILING DATE: 2000-06-19

PRIOR APPLICATION NUMBER: 60/212,681

PRIOR FILING DATE: 2000-06-20

PRIOR APPLICATION NUMBER: 60/290,170

PRIOR FILING DATE: 2001-05-10

PRIOR APPLICATION NUMBER: PA 2001 00739

PRIOR FILING DATE: 2001-05-10

NUMBER OF SEQ ID NOS: 24

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 4

LENGTH: 239

TYPE: PR1

ORGANISM: Aequorea victoria

US-10-296-953-4

Query Match 99.7%; Score 1269; DB 12; Length 239;

Best Local Similarity 99.6%; Pred. No. 5.8e-124;

Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNHHKFSVSGEGDATYKLTLPICITTKGLPVPWPT 60

DB 1 MVSKEELFTGVVPILVELDGVNHHKFSVSGEGDATYKLTLPICITTKGLPVPWPT 60

QY 61 LVTTLSGVGVCFSRYPDHMKQHOFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120

DB 61 LVTTLSGVGVCFSRYPDHMKQHOFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 131 DHYQONTPIGDGPVLLPDNHYLTSQSALSCKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

DB 131 DHYQONTPIGDGPVLLPDNHYLTSQSALSCKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 3

US-10-270-223-6

Sequence 5, Application US/10270223

Publication No. US20030143634A1

GENERAL INFORMATION:

APPLICANT: BioImage A/S

TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR COMPONENTS

TITLE OF INVENTION: INTERACT LIVING CELLS, AND TO EXTRACT QUANTITATIVE INFORMATION RE

TITLE OF INVENTION: INTERACTIONS BY FLUORESCENCE REDISTRIBUTION.

FILE REFERENCE: 3759-0126P

CURRENT APPLICATION NUMBER: US/10/270,223

CURRENT FILING DATE: 2002-10-11

NUMBER OF SEQ ID NOS: 12

SOFTWARE: PatentIn version 3.1

SEQ ID NO 6

LENGTH: 363

TYPE: PR1

ORGANISM: Aequorea Victoria and Human

US-10-270-223-6

Query Match 99.7%; Score 1269; DB 14; Length 363;

Best Local Similarity 99.6%; Pred. No. 1.1e-123;

Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
; SEQ ID NO 32
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-32

Query Match          99.7%; Score 1269; DB 14; Length 1132;
Best Local Similarity 99.6%; Pred. No. 5.4e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 894 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTFLKFICTTGKLPVPWPT 953
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 954 LVTLSYGVCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 1013
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180.
DB 1014 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 1073
QY 181 DHYQQNTPIGDGPVLLPDNNHYLSTQALSQDPNEKRDHMLGFTVTAAGITLGMDELYK 239
DB 1074 DHYQQNTPIGDGPVLLPDNNHYLSTQALSQDPNEKRDHMLGFTVTAAGITLGMDELYK 1132

RESULT 6
US-09-887-784-2
; Sequence 2, Application US/09887784
; Patent No. US20020177189A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea Victoria
US-09-887-784-2

Query Match          99.1%; Score 1261; DB 9; Length 239;
Best Local Similarity 99.2%; Pred. No. 4e-123;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTFLKFICTTGKLPVPWPT 60
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLSYGVCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQQNTPIGDGPVLLPDNNHYLSTQALSQDPNEKRDHMLGFTVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGPVLLPDNNHYLSTQALSQDPNEKRDHMLGFTVTAAGITLGMDELYK 239

RESULT 7
US-10-296-953-2
; Sequence 2, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match          98.8%; Score 1258; DB 9; Length 239;
Best Local Similarity 98.7%; Pred. No. 8.1e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTFLKFICTTGKLPVPWPT 60
```

```
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match          99.1%; Score 1261; DB 12; Length 239;
Best Local Similarity 99.2%; Pred. No. 4e-123;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTFLKFICTTGKLPVPWPT 60
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLSYGVCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQQNTPIGDGPVLLPDNNHYLSTQALSQDPNEKRDHMLGFTVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGPVLLPDNNHYLSTQALSQDPNEKRDHMLGFTVTAAGITLGMDELYK 239

RESULT 8
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. US20020083488A1
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match          98.8%; Score 1258; DB 9; Length 239;
Best Local Similarity 98.7%; Pred. No. 8.1e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTFLKFICTTGKLPVPWPT 60
```





APPLICANT: TSJEN, Roger  
APPLICANT: ZACHARIAS, David  
APPLICANT: BAIRD, Geoffrey  
TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS  
FILE REFERENCE: REGEN1530  
CURRENT APPLICATION NUMBER: US/09/794,308  
CURRENT FILING DATE: 2001-02-26  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
US-09-794-308-4

Query Match 98.8%; Score 1258; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 8.1e-123;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYKLTAKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYKLTAKFICTTGKLPVPWPT 60  
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 13  
US-09-865-291-4  
Sequence 4, Application US/09865291  
Publication No. US20030186229A1  
GENERAL INFORMATION:  
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA  
APPLICANT: TSJEN, Roger  
APPLICANT: TING, Alice  
APPLICANT: ZHANG, Jin  
TITLE OF INVENTION: EMISSION RATIO-METRIC INDICATORS OF PHOSPHORYLATION  
FILE REFERENCE: REGEN1550  
CURRENT APPLICATION NUMBER: US/09/865,291  
CURRENT FILING DATE: 2001-05-24  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
US-09-865-291-4

Query Match 98.8%; Score 1258; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 8.1e-123;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYKLTAKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYKLTAKFICTTGKLPVPWPT 60  
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 14  
US-10-457-982-3  
Sequence 3, Application US/10457982  
Publication No. US20030212265A1  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger Y.  
APPLICANT: Miyawaki, Atsushi  
APPLICANT: Llopis, Juan  
APPLICANT: Wachter, Rebekka M.  
APPLICANT: Remington, S. James  
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
FILE REFERENCE: 07257/071001  
CURRENT APPLICATION NUMBER: US/10/457,982  
CURRENT FILING DATE: 2003-06-09  
PRIOR APPLICATION NUMBER: US/09/602,641  
PRIOR FILING DATE: 2000-06-22  
PRIOR APPLICATION NUMBER: 09/172,063  
PRIOR FILING DATE: 1998-10-13  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (0)...(0)  
OTHER INFORMATION: EGFP  
US-10-457-982-3

Query Match 98.8%; Score 1258; DB 12; Length 239;  
Best Local Similarity 98.7%; Pred. No. 8.1e-123;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYKLTAKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYKLTAKFICTTGKLPVPWPT 60  
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 15  
US-10-121-258-13  
Sequence 13, Application US/10121258  
Publication No. US20030059835A1  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger  
APPLICANT: Campbell, Robert  
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT  
TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME  
FILE REFERENCE: UC083.1CP2CP1  
CURRENT APPLICATION NUMBER: US/10/121,258  
CURRENT FILING DATE: 2002-04-10  
PRIOR APPLICATION NUMBER: 09/794,308  
PRIOR FILING DATE: 2001-02-26  
PRIOR APPLICATION NUMBER: 09/866,538  
PRIOR FILING DATE: 2001-05-24

Query Match 98.8%; Score 1258; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 8.1e-123;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYKLTAKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYKLTAKFICTTGKLPVPWPT 60  
QY 61 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

```
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Enhanced Green Fluorescent Protein (EGFP)
US-10-121-258-13

Query Match      98.8%; Score 1258; DB 14; Length 239;
Best Local Similarity 98.7%; Pred. NO. 8.1e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSSEGECDATYGLTKLFICTTGKLPVPWPT 60
   |||||
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSSEGECDATYGLTKLFICTTGKLPVPWPT 60
   |||||

QY 51 LVTALSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
   |||||
Db 51 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
   |||||

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
   |||||
Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
   |||||

QY 131 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLGFVTAAGITLGMDELYK 239
   |||||
Db 131 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLGFVTAAGITLGMDELYK 239
   |||||

Search completed: June 21, 2004, 16:09:26
Job time : 35.7778 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:00 ; Search time 10.3333 Seconds  
(without alignments)  
2224.817 Million cell updates/sec

Title: US-09-887-784-64A  
Perfect score: 1273  
Sequence: 1 MVSKGEELFTGVVPILVELD.....VLLGFVTAAGITLGMDELYK 239  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1236	97.1	238	1 JQ1514	green-fluorescent
2	105	8.2	785	2 H72228	hypothetical prote
3	92.5	7.3	655	2 D83917	DNA topoisomerase
4	91.5	7.2	370	2 E70390	iron-sulfur cofact
5	91.5	7.2	861	2 H64102	leucine-tRNA ligas
6	89.5	7.0	887	2 E82590	leucyl-tRNA synthe
7	89	7.0	439	2 JH0414	synaptogamin o-p65
8	87.5	6.9	860	2 AC0582	leucyl-tRNA synthe
9	87.5	6.9	2573	2 D71614	hypothetical prote
10	87	6.8	578	1 I40794	dihydrolipoamide d
11	86.5	6.8	797	2 JC4078	protective surface
12	86.5	6.8	808	2 F64102	cellulase (EC 3.2.
13	86.5	6.8	941	2 S29043	cellulase (EC 3.2.
14	86	6.8	357	2 G81355	tRNA (uracil-5)-m
15	85.5	6.7	788	1 JDLVHH	DNA-directed DNA p
16	85.5	6.7	889	2 JC5576	inter-alpha-trypsi
17	85	6.7	281	2 AD2052	hypothetical prote
18	85	6.7	632	2 T06586	DNA-binding protei
19	84.5	6.6	425	2 C97354	hypothetical prote
20	84.5	6.6	613	2 A99552	oligodeoxyphosphat
21	84.5	6.6	740	2 G95153	neuraminidase, pro
22	83.5	6.6	836	1 JDLVD	DNA-directed DNA p
23	83.5	6.6	1134	2 A60234	IgA Fc receptor pr
24	83.5	6.6	1164	1 FCSOAG	IgA Fc receptor pr
25	83	6.5	461	2 T06936	photosystem II chl
26	83	6.5	471	2 T27856	hypothetical prote
27	83	6.5	774	2 T39539	alpha-amylase homo
28	82.5	6.5	2222	1 A36028	DNA-directed DNA p
29	82	6.4	353	2 E84941	imidazoleglycerol-

ALIGNMENTS

RESULT 1

JQ1514  
green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)  
C:Species: Aequorea victoria  
C:Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 23-Mar-2001  
C:Accession: JQ1514; PQ0335; S48693; S51330; S51331  
R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.  
Gene 111, 229-233, 1992  
A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.  
A:Reference number: JQ1514; MUID:92175527; PMID:1347277  
A:Accession: JQ0692  
A:Molecule type: DNA  
A:Residues: 1-107, 'S', 109-238 <PRA1>  
A:Cross-references: GB:M62654; NID:G155662; PIDN:AAA27722.1; PID:G155663  
A:Accession: JQ1514  
A:Molecule type: mRNA  
A:Residues: 1-99, 'F', 101-140, 'L', 142-218, 'V', 220-238 <PRA2>  
A:Cross-references: GB:M62653; NID:G155660; PIDN:AAA27721.1; PID:G155661  
A:Accession: PQ0335  
A:Molecule type: protein  
A:Residues: 46-64; 74-122; 132-151; 154-183; 185-200 <PRA3>  
R:Inouye, S.; Tsuji, F.I.  
FEBS Lett. 351, 211-214, 1994  
A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.  
A:Reference number: S48693; MUID:94364470; PMID:8082767  
A:Accession: S48693  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-24, 'O', 26-156, 'P', 158-171, 'K', 173-238 <INO>  
A:Cross-references: GB:L29345; NID:G606383; PIDN:AAA58246.1; PID:G606384  
R:Watkins, J.N.; Campbell, A.K.  
Submitted to the EMBL data Library, January 1995  
A:Reference number: S51330  
A:Accession: S51330  
A:Molecule type: mRNA  
A:Residues: 1-13, 'V', 15-24, 'O', 26-44, 'N', 46-153, 'G', 155-156, 'P', 158-171, 'K', 173-227, 'R', 229-238 <PRA4>  
A:Cross-references: EMBL:X83959; NID:G634008; PIDN:CAA58789.1; PID:G634009  
A:Experimental source: clone gfp1  
A:Accession: S51331  
A:Molecule type: mRNA  
A:Residues: 1-24, 'O', 26-29, 'R', 31-83, 'L', 85-153, 'G', 155-156, 'P', 158-171, 'K', 173-208, 'Q', 209-238 <PRA5>  
A:Cross-references: EMBL:X83960; NID:G634010; PIDN:CAA58790.1; PID:G634011  
A:Experimental source: clone gfp2  
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.  
Submitted to the Brookhaven Protein Data Bank, August 1996  
A:Reference number: A65692; PDB:IGFL  
A:Contents: annotation; X-ray crystallography, 1.9 angstroms, residues 'A', 2-79, 'R', 81-91; alpha-amylase homo  
A:Note: engineered sequence based on JQ1514, cloned and expressed in Escherichia coli  
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.  
Nat. Biotechnol. 14, 1246-1251, 1996  
A:Title: The molecular structure of green fluorescent protein.  
A:Reference number: A58953; MUID:98294543; PMID:9631087

hypothetical prote  
hypothetical prote  
carboxy-terminal p  
proprotein convert  
protein F10B6.14  
lipoxigenase (EC 1  
water-stress-induc  
enkephalin precurs  
synegobymenotropi  
polyketide synthas  
hypothetical prote  
hypothetical prote  
hypothetical prote  
coatomer complex a

A:Content: annotation; X-ray crystallography, 1.9 angstroms  
C:Comment: This protein is excited by the photoprotein aequorin (see PIR-AQJFNV) emitting  
C:Comment: the chromophore of this protein is formed by modification of Ser-dehydro-Tyr-  
C:Genetics:

A:Gene: GFP  
A:Introns: 69/3; 167/3  
C:Superfamily: green-fluorescent protein  
C:Keywords: chromoprotein; luminescence  
F:55-67/Cross-link: 5-imidazolinone (Ser-Gly) #status experimental  
F:66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match 97.1%; Score 1236; DB 1; Length 238;  
Best Local Similarity 96.6%; Pred. No. 2.1e-96;  
Matches 230; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIILVELDGVNKHFSVSGEGDATYKGLTKFICTTGKLPVWPPTL 61  
DB 1 MSKGEELFTGVVPIILVELDGVNKHFSVSGEGDATYKGLTKFICTTGKLPVWPPTL 60  
QY 62 VTALSVCVQCFSRYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNKYKTRAEVKFEGDTLV 121  
DB 61 VTFYSVCVQCFSRYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNKYKTRAEVKFEGDTLV 120  
QY 122 NRIELKGIDFKEDGNILGHKLEYNVSHVYIMADKQNGIKVNFIRHNIEDGSVOLAD 181  
DB 121 NRIELKGIDFKEDGNILGHKLEYNVSHVYIMADKQNGIKVNFIRHNIEDGSVOLAD 180  
QY 182 HYQONTPIGDPVLLPDNHYLSQSALSKDPNEKRDMHVLGFTVTAAGITLGMDELYK 239  
DB 181 HYQONTPIGDPVLLPDNHYLSQSALSKDPNEKRDMHVLGFTVTAAGITLGMDELYK 238

## RESULT 2

H72228  
hypothetical protein TW1624 - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: H72228  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.

Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; UID:99287316; PMID:10360571  
A:Accession: H72228

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-785 <ARN>  
A:Cross-references: GB:AE001806; GB:AE000512; NID:g4982196; PIDN:AAD36691.1; PID:g498219

A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TW1624

Query Match 8.2%; Score 105; DB 2; Length 785;  
Best Local Similarity 19.7%; Pred. No. 0.86;  
Matches 46; Conservative 32; Mismatches 71; Indels 84; Gaps 7;

QY 3 SKGEELFTGVVPIILVELDGVNKHFSVSGEGDATYKGLTKFICTTGKLPVWPPTLV 62  
DB 15 NEGRFSEGTVPGVQAD-----LVRKGLLPHYPVGM- 46  
QY 63 TALSVCVQCFSRYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNKYKTRAEVKFEGDTLV 122  
DB 47 -----NEDLFKEIDREWIYEREFKEDVKEGEDVLDFEGVDTL 88  
QY 123 RIELKGIDFKEDGNILGHKLEYNVSHVYIMADKQNGIKVNFIRHNIEDGSVOLADH 182  
DB 89 DVLNGVYL--GSTDMEFYRFDVTNL-----KEKNHLKVYIK-----SPIRVPKT 134  
QY 133 YQONTPIGDPVLLPDNHYLSQSALSKDPNEKRDMHVLGFTVTAAGITLGM 235  
DB 135 LEQNYGLVGGP-----EDP-----IRGYIRKQAYSYGWD 163

## RESULT 3

DB3917  
DNA topoisomerase IV subunit B BH2140 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hiran  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: AB3650; UID:20512582; PMID:11058132  
A:Accession: DB3917  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-655 <STO>  
A:Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BA005859.1; GSPDB:GN001

A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH2140  
C:Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B

Query Match 7.3%; Score 92.5; DB 2; Length 655;  
Best Local Similarity 22.3%; Pred. No. 7.6;  
Matches 54; Conservative 37; Mismatches 94; Indels 57; Gaps 9;

QY 22 DVNGHK---FSVSGEGEGDAT---YGKLTLLKFI-----CTTGKLPVWP 59  
DB 63 NVTIHKDQSVVRDEGRGMPGTGMHKLKGTPEVILTVLHAGGKFGGGYATSGGLHGVGA 122  
QY 60 TLVTALSVCVQCFSRYPDHMKQHDFFKSAMPEGYVOER-----TIFKDDG----- 105  
DB 123 SVVNALSEWLIVEIKRDGMVYEQRFENGKGPSTLLEKKGKTRGTTHFKPDPTVFSTT 182  
QY 106 --NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNVSHVYIMADK----- 157  
DB 183 NFNVTLSRLREAAFLKGLKIELVDLRDRTKEVPH-YEDGKAKFVYLNEDKETHLPV 241  
QY 158 -----QKNGIKVNFIRHNIEDGSVOLADHYQONTPIGDPVLLPDNHYLSQSALSKDP 212  
DB 242 VFPENGSGNIEIEFAPOFN--DGYTENVLSFVNNVTRTKDG-----GTHELGAKTAMTRAV 294  
QY 213 NE 214  
DB 295 NE 296

## RESULT 4

E70390  
iron-sulfur cofactor synthesis protein nifs - Aquifex aeolicus  
N:Contains: L-cysteine sulfotransferase (EC 2.8.1.-)  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 07-Dec-1999

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Over  
V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; UID:98196666; PMID:9537320  
A:Accession: E70390  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-370 <AQF>  
A:Cross-references: GB:AE000720; NID:g2983529; PIDN:AAC07111.1; PID:g2983536; GB:AE000657.

A:Experimental source: strain VF5  
C:Genetics:  
A:Gene: nifs1  
C:Superfamily: nitrogen fixation protein nifs  
C:Keywords: phosphoprotein; pyridoxal phosphate; sulfotransferase  
F:195/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted  
F:318/Active site: Cys (cysteine persulfide intermediate) #status predicted

Query Match 7.2%; Score 91.5; DB 2; Length 370;  
Best Local Similarity 25.9%; Pred. No. 4.4;

Matches 50; Conservative 30; Mismatches 90; Indels 23; Gaps 8;  
QY 4 KGBELFTGVV-----PILVELD---GDVNGHKF-SVSGEG-----EGDATYGKLTLPFICT 50  
Db 164 KGVPLLTDAVQAGKPIELKNISYATFSGKKHAKGSGFLYSDEANVEPLIVGGQE 223  
QY 51 TGKLP-----VPMPTLVTALSYGQCFSRYPDHMKQ-HDFPKSAMPEGYVQERTIFFPKDD 104  
Db 224 NGKRSGTENVVGLISLAKALEIIVNFSRYOEQLKRLDLFENLLLEA-LPDAQIVGKDA 282  
QY 105 GNYKTRAEV---KFEQDGLVNRLELGDIDPKEDGNILGHLEYNVSHNVYIMADKQKNG 161  
Db 283 ERSPSISVVIMPKFPAEIVNKLSEGIYCTSGSACLSEGEYEPNKMVKMGFSQEKALRM 342  
QY 162 IKVNFKIRHNIED 174  
Db 343 VRESFGLLNKEE 355  
RESULT 5  
H64102  
leucine-tRNA ligase (EC 6.1.1.4) - Haemophilus influenzae (strain Rd KW20)  
N/Alternate names: leucyl-tRNA synthetase  
C/Species: Haemophilus influenzae  
C/Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 03-Jun-2002  
C/Accession: H64102  
R/Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.  
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,  
A./Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A/Reference number: A64000; MUID:95350630; PMID:7542800  
A/Accession: H64102  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-861 <TIGR>  
A/Cross-references: GB:U32774; GB:L42023; NID:g1573942; PIDN:AAC22581.1; PID:g1573943;  
C/Genetics:  
A/Gene: leuS  
C/Suprafamily: leucine-tRNA ligase  
C/Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis  
Query Match 7.2%; Score 91.5; DB 2; Length 861;  
Best Local Similarity 24.1%; Pred. No. 13;  
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;  
QY 50 TTGKLPVPMPTLVTALSYGQCFSRYPDHMKQHDFFKSAAMPEGYVQERTIFFPKD----- 103  
Db 314 TGDKLPi-WVANFVLMHYGTGAVMAVPAH-QDRDF-----EFAQKYSLPiKQVIAPLA 364  
QY 104 DGNKTRAEVKPGDGLVNRLELKGIDFKEDGNILGHLEYNVSHNVYIMADK-QKNGI 162  
Db 365 DESIDITKQAFVEHGLKLVNSDEFDGKNF--DGAENG-----IADKLEKLV 408  
QY 163 ---KVNFKIRH-----NIEDGVSQVLADHYVQNTPIGDGVLPLPDNHYL- 202  
Db 409 GKQVNYRLDGVSRQRYWGAPIPMLTLENGDVVPA-----PMEDPILPPEVDVMD 461  
QY 203 STQSALSADPN 213  
Db 462 GVKSPINADPN 472  
RESULT 6  
E82590  
leucyl-tRNA synthetase xf2176 [imported] - Xylella fastidiosa (strain 9a5c)  
C/Species: Xylella fastidiosa  
C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
C/Accession: E82590  
R/anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
Nature 406, 151-157, 2000  
A/Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A/Reference number: A82515; MUID:20365717; PMID:10910347  
A/Note: for a complete list of authors see reference number A59328 below  
A/Accession: E82590  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-887 <SIM>  
A/Cross-references: GB:AE004031; GB:AE003849; NID:g9107309; PIDN:AAF84975.1; GSPDB:GN001  
A/Experimental source: strain 9a5c  
R/Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Laig  
Briões, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Canargo, L.E.A.; Carraro, D.M.; Carrer, H  
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig  
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
A/Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.  
; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A/Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A/Reference number: A59328  
A/Contents: annotation  
C/Genetics:  
A/Gene: XF2176  
C/Suprafamily: leucine-tRNA ligase  
Query Match 7.0%; Score 89.5; DB 2; Length 887;  
Best Local Similarity 22.2%; Pred. No. 20;  
Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;  
QY 50 TTGKLPVPMPTLVTALSYGQCFSRYPDHMKQHDFFKSAAMPEGYVQERTIFFKDDGNY-- 107  
Db 329 TNEQLFV-WVANFVLMAYGTGAVMAVPGHDQRQDEF--ANKYGLPIRQVIALKEPKNQDE 385  
QY 108 -----KTRAEVKFGDGLVNRLELKGIDFKEDGNILGHLEYNVSHNVYI 153  
Db 386 STWEPDVRWDYADKTR---EPE---LINSAPDGLDYQDAFEVLAERF----- 429  
QY 154 MADKQKNG-IKVNFKIRHNIEDGVSQVLADHYVQNTPI-----GDGPVLLPDN 199  
Db 430 ---RQGRQRYVNYRLR---DWGVSQRVWGCPiVIVYPTCGAVPVPEDQPLVILPEN 482  
QY 200 -HYLSTQSALSADPN 216  
Db 483 VAFSGTSGPIKTDPEWRK 500  
RESULT 7  
JH0414  
synaptogamin o-p65-B - electric ray (Discothyrea ommata)  
N/Alternate names: synaptic vesicle protein o-p65-B  
C/Species: Discothyrea ommata  
C/Date: 15-Jan-1993 #sequence\_revision 15-Jan-1993 #text\_change 13-Aug-1999  
C/Accession: JH0414; PS0223  
R/Wendland, B.; Miller, K.G.; Schilling, J.; Scheller, R.H.  
Neuron 6, 993-1007, 1991  
A/Title: Differential expression of the p65 gene family.  
A/Reference number: JH0413; MUID:91273991; PMID:2054189  
A/Accession: JH0414  
A/Molecule type: mRNA  
A/Residues: 1-439 <WEN>  
A/Cross-references: GB:M64276; NID:g213110; PIDN:AAA49228.1; PID:g213111  
A/Experimental source: electric organ  
A/Accession: PS0223  
A/Molecule type: protein  
A/Residues: 'MLV', '26-34', 'XX', '194-199', 'X', '201-206', 'X', '322-332', 'D', '334-337' <WEN1>  
C/Suprafamily: synaptotagmin; protein kinase C C2 region homology  
C/Keywords: glycoprotein; membrane protein; synaptic vesicle  
F/75-101/Domain: hydrophobic <HYD>  
F/153-266/Domain: protein kinase C C2 region homology <KC2A>  
F/284-399/Domain: protein kinase C C2 region homology <KC2B>  
F/6,46/Binding site: carbohydrate (Asn) (covalent) #status predicted

```
Query Match 7.0%; Score 89; DB 2; Length 439;
Best Local Similarity 19.5%; Pred. No. 8.9;
Matches 56; Conservative 46; Mismatches 93; Indels 92; Gaps 12;

QY 56 LVLELDGVDNGHFKFSVSGEGEGDATYKGLTLKFCTTGKLPV-PWPTLVLTALSYGV----- 69
DB 38 MNPIDTGDNSTEAGVPGEGKND-VPEKLEKFMNELQKIPLPWALIAIIVSGLLLTLC 96
QY 70 -----QCSRYDPDHMKQHDFFKSAPEGVQVQRTIFFKDDGNYKTRAEVKPEG----- 117
DB 97 CLCICKKCCCKKKKKKKGK-----KNDINMK---DVKGSGGNQDD 138
QY 118 --DTLVNRTELKIDFKEDGNI--LGHKLEYNNYNSH----- 149
DB 139 DAETGTEGDEKEEAKKEKLGKIOFSLDYDPOANQLTVGIQAAELPALDMGTSDFY 198
QY 150 -NYIYMADKQKN-GIKVN-----FKIRHNIEDGSVQLA-----DHYQOQTFI 189
DB 199 VKVFLLPDKKKYETVQKTLNPTFNESFVKVPYQELGGKTLMAVYDFDRFSKHDCI 258
QY 190 GDGPVLLPD-----NHVLSQSALSQKDPNEKRDMVLLGFVTAAG 229
DB 259 GQTVLMTKVDLQQLQLEWRDLESABKEPEKLGIDICTSLRYVPTAG 305

RESULT 8
AC0582
A:Title: leucyl-tRNA synthetase [imported] - Salmonella enterica subsp. enterica serovar Typhi (S
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AC0582
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
C:Accession: AC0582
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-860 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD05125.1; PID:G16501899; GSPDB:GN00176
C:Genetics:
A:Gene: STY0699
C:Superfamily: leucine-tRNA ligase

Query Match 6.9%; Score 87.5; DB 2; Length 860;
Best Local Similarity 23.3%; Pred. No. 29;
Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;

QY 50 TTGKLPVWPPTLVLTALSYGVQCSRYDPDHMKQHDFFKSAPEGVQVQRTIFFKDDGNYKT 109
DB 314 TGEIPIV-WAANFVMEYGTGAVMVPGH-DQD-YEFASKYGLTIKPVILAADGSEPD 370
QY 110 RAEVKFEGDTLVNRIELKIDGDFKEDGNILGHKLEYNNYNSHVMADKQNGIKVNFKIR 169
DB 371 SEQALTEKGVLFNSGFDFGLAFEAFAAFAIAADKL-----AEKGVGERKVMYRLR 418
QY 170 H-----NIEDGSVOLADHYQOQTFPIGDGPVLLPDNHYL-STQSALSQDP 212
DB 419 DWGVSQRVYGAIPWVLTLEDGV-----LTFPDQLPVILPVEDVMDGIISPICADP 471

RESULT 9
D71614
hypothetical protein PF0460c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: D71614
R:Gardner, M.J.; Tetelin, H.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
```

```
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: D71614
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2573 <GAR>
A:Cross-references: GB:AE001396; GB:AE001388; NID:G3845188; PIDN:AAC71881.1; PID:G384519
A:Experimental source: Clone 3D7
C:Genetics:
A:Gene: PF0460c

Query Match 6.9%; Score 87.5; DB 2; Length 2573;
Best Local Similarity 26.2%; Pred. No. 1.2e+02;
Matches 34; Conservative 30; Mismatches 53; Indels 13; Gaps 5;

QY 94 VOERTIFFKD--DGNVYKTRAEVKFEGDTLVNRIELKIDGDFKEDGNILGHKLEYN--YNH 149
DB 126 LKKTETILCKDIKXSGSDPMDDEISLFPKDDMVDDKELK--DFEKSLSKIKNKEVNFYNNKM 183
QY 150 NYIYMADKQNGIKVNFKIRHNIEDGSVOLADHYQOQTFPIGDGPVLLPDNHYLSTQSALS 209
DB 184 NLHIEKNKKKDEKKNKIHNNDENNM-----IYYKNI---DKTHYLDNNVVHILNDIN 236
QY 210 KDPNEKRDM 219
DB 237 TVLKRERYM 246

RESULT 10
I40794
dihydropyrimidine dehydrogenase (EC 1.8.1.4) [validated] - Clostridium magnum
N:Alternate names: 2-oxoglutarate dehydrogenase complex chain E3; acetoin dehydrogenase
hydrogenase complex chain E3; S-complex 50K chain
C:Species: Clostridium magnum
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: I40794
R:Kruger, N.; Oppermann, F.B.; Lorenzl, H.; Steinbuechel, A.
J. Bacteriol. 176, 3614-3630, 1994
A:Title: Biochemical and molecular characterization of the Clostridium magnum acetoin de
A:Reference number: I40789; MUID:94266715; PMID:8206840
A:Accession: I40794
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-578 <KRU>
A:Cross-references: GB:I31844; NID:G472324; PIDN:AAA21748.1; PID:G472330
C:Function:
A:Description: catalyzes the oxidation of dihydropyrimidine to lipamide using NAD
A:Pathway: acetoin dehydrogenase enzyme system
C:Superfamily: Alkaligenes dihydropyrimidine dehydrogenase; dihydropyrimidine dehydrogenase
C:Keywords: FAD; flavoprotein; lipamide; NAD; oxidoreductase; redox-active disulfide
F;S-77/Domain: lipoyl/biotin-binding homology <LPS>
F;117-145/Region: beta-alpha-beta FAD nucleotide-binding fold
F;119-561/Domain: dihydropyrimidine dehydrogenase homology <DLD>
F;287-315/Region: beta-alpha-beta NAD nucleotide-binding fold
F;153-158/Disulfide bonds: redox-active #status predicted

Query Match 6.8%; Score 87; DB 1; Length 578;
Best Local Similarity 23.8%; Pred. No. 19;
Matches 54; Conservative 40; Mismatches 84; Indels 52; Gaps 12;

QY 10 TGVVPIVLVDGVDNGHFKFSVSGEGEGDATYKGLTLK-----FICTTGKLPVWPPTLV 63
DB 255 TGSMPPIPIE-----GNKLS---GVIDST-GALSLENPESITAIIGGVIGVEFASIFN 305
QY 64 ALSYGVQCSRYDPDHMKQHDFFKSAPEGVQVQRTIFFKDDGNYKTRAEVKFEGDTLVNR 123
DB 306 SLGCKVSIIEMLPHILPMDREISEI-----AKAKLIRDGINNN 346
QY 124 IELKIDGDFKEDG---NILGHKLEYNNYNSHVMADKQKN--GIKNFKIRHNIEDGSVQ 178
DB 347 CKVTRIEQGEDGLKVSFIDGKGESIDVEKVLIAVGRRSNIEGLDVE-KLGKTEGSGII 405
```

[illegible]

RESULT 14		QY	203	STQ-----SALS	KDPNEKR 216
G81355		Db	298	DRRRQKGGVQLAISREPS	ETR 320
tRNA (uracil:5)-methyltransferase (EC 2.1.1.35) Cj0831c [imported] - Campylobacter jejuni		Search completed: June 21, 2004, 16:01:55			
C:Species: Campylobacter jejuni		Job time : 11.3333 secs			
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002					
C:Accession: G81355					
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling					
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell					
Nature 403, 665-668, 2000					
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp					
A:Reference number: A81250; MUID:20150912; PMID:10688204					
A:Accession: G81355					
A>Status: preliminary					
A:Molecule type: DNA					
A:Residues: 1-357 <PAR>					
A:Cross-references: GB:AL139076; GB:AL111168; NID:G9698128; PIDN:CAB73096.1; PID:G969827					
A:Experimental source: serotype O2, strain NCTC 11168					
C:Genetics:					
A:Gene: tmaA; Cj0831c					
C:Keywords: methyltransferase; S-adenosylmethionine					
Query Match		6.8%; Score 86; DB 2; Length 357;			
Best Local Similarity		24.8%; Pred. No. 12;			
Matches		30; Conservative 18; Mismatches 39; Indels 34; Gaps 5;			
QY	30	KQHDFPKSAMPEGYQVQERTIFFKDDGNYKTRAEVKF--EGDTLV-----NRIELKG 128			
Db	14	EKHSFIKKYFKFYTKDFKLFAKDKHYRTRAELSFYHENDTLFYAMFDPKSKKKYIIIEY 73			
QY	129	IDFKED-----GNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIE 173			
Db	74	LDFADEKICAFMPRLLELYLRQDNKLEKL-----FGVEFLTTKQE--LSITLLYHKNIE 125			
QY	174	D 174			
Db	126	D 126			
RESULT 15					
JDVLHH					
DNA-directed DNA polymerase (EC 2.7.7.7) - heron hepatitis virus					
C:Species: heron hepatitis virus, HHBV					
A:Note: host Ardea cinerea (Gray heron)					
C>Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 11-Jun-1999					
C:Accession: A30082					
R:Sprenkel, R.; Kaleta, E.F.; Will, H.					
J. Virol. 62, 3832-3839, 1988					
A:Title: Isolation and characterization of a hepatitis B virus endemic in herons.					
A:Reference number: A93037; MUID:88333160; PMID:3418788					
A:Accession: A30082					
A:Molecule type: DNA					
A:Residues: 1-788 <SPR>					
A:Cross-references: GB:M22056; NID:G325452; PIDN:AAA45738.1; PID:G325454					
C:Superfamily: hepatitis virus DNA-directed DNA polymerase					
C:Keywords: DNA biosynthesis; nucleotidyltransferase					
Query Match		6.7%; Score 85.5; DB 1; Length 788;			
Best Local Similarity		19.7%; Pred. No. 38;			
Matches		40; Conservative 32; Mismatches 66; Indels 65; Gaps 10;			
QY	58	WPTLVLTALSYGVCFSRYPDHMKQH-----DFFKSAMPEGYQVQERT----IFFKDDGNYK 108			
Db	139	WPKSISYLVPHSGVGVKPKYPEQQNHESLVNDYLNKLFEGAGILYKRVSKHLVTFK--GPYF 196			
QY	109	T-----RAEVKFEQDTLVNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIM 154			
Db	197	TWEQKHLVPQOHGAYGSKINDRQESRRRIITATSRKNDSSRI-----FGAHN----- 245			
QY	155	ADKQKNGIKVNFKIRHNIEGDSVOLADHYQ-----QNTPIGDGPVLL--PDNHYL 202			
Db	246	-----NGRKISY---HSTRDGSRLSGRTSDPTSRGALAGGDSPTPGGTAHPSTHHV 297			



RESULT 1

ID	GFP_AEQVI	STANDARD;	PRT;	238 AA.
AC	P42212; Q17104; Q27903;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Green fluorescent protein.			
DE	GFP.			
OS	Aequorea victoria (Jellyfish).			
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;			
OC	Aequoreidae; Aequorea.			
OX	NCBI_TaxID=6100;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=92175527; PubMed=1347277;			
RA	Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,			
RA	Cormier M.J.;			
RT	"Primary structure of the Aequorea victoria green-fluorescent			
RT	protein.";			
RL	Gene 111:229-233(1992).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94185810; PubMed=8137953;			
RA	Inouye S., Tsuji F.I.;			
RT	"Aequorea green fluorescent protein. Expression of the gene and			
RT	fluorescence characteristics of the recombinant protein.";			
RL	FEBS Lett. 341:277-280(1994).			
RL	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97299832; PubMed=9154981;			
RA	Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;			
RT	"Enhanced expression in tobacco of the gene encoding green fluorescent			
RT	protein by modification of its codon usage.";			
RL	Plant Mol. Biol. 33:989-999(1997).			
RL	[4]			
RP	CHROMOPHORE.			
RX	MEDLINE=93192221; PubMed=8448132;			
RA	Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;			
RT	"Chemical structure of the hexapeptide chromophore of the Aequorea			
RT	green-fluorescent protein.";			
RL	Biochemistry 32:1212-1218(1993).			
RL	[5]			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=96355665; PubMed=8703075;			
RA	Ormoie M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,			
RA	Remington S.J.;			
RT	"Crystal structure of the Aequorea victoria green fluorescent			
RT	protein.";			
RL	Science 273:1392-1395(1996).			
RL	[6]			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=98294543; PubMed=9631087;			
RA	Yang F., Moss L.G., Phillips G.N. Jr.;			
RT	"The molecular structure of green fluorescent protein.";			
RL	Nat. Biotechnol. 14:1246-1251(1996).			

[7]  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.  
RX MEDLINE=9845509; PubMed=9782051;  
RA Wachter R.M., Elsiger M.A., Kallio K., Hanson G.T., Remington S.J.;  
RT "Structural basis of spectral shifts in the yellow-emission variants  
of green fluorescent protein.";  
RL Structure 6:1267-1277(1998).  
[8]  
RN X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
RX MEDLINE=99238303; PubMed=10220315;  
RA Elsiger M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J.;  
RT "Structural and spectral response of green fluorescent protein  
variants to changes in pH";  
RL Biochemistry 38:5296-5301(1999).  
CC -!- FUNCTION: Energy-transfer acceptor. Its role is to transduce the  
blue chemiluminescence of the protein aequorin into green  
fluorescent light by energy transfer. Fluoresces in vivo upon  
receiving energy from the Ca(2+)-activated photoprotein aequorin.  
CC Absorbs light maximally at 395 nm and exhibits a smaller  
absorbance peak at 470 nm. The fluorescence emission spectrum  
peaks at 509 nm with a shoulder at 540 nm.  
CC -!- SUBUNIT: Monomer.  
CC -!- TISSUE SPECIFICITY: Photocytes.  
CC -!- PTM: Contains a covalently attached chromophore, which is composed  
of modified amino acid residues. The chromophore is formed upon  
cyclization of the residues Ser-dehydroTyr-Gly.  
CC -!- BIOTECHNOLOGY: Has become a useful and ubiquitous tool for making  
chimeric proteins of GFP linked to other proteins where it  
functions as a fluorescent protein tag. GFP tolerates N- and C-  
terminal fusion to a broad variety of proteins. It has been  
expressed in bacteria, yeast, slime mold, plants, Drosophila,  
zebrafish, and in mammalian cells. As a noninvasive fluorescent  
marker in living cells, it allows for a wide range of applications  
where it may function as a cell lineage tracer, reporter of gene  
expression, or as a measure of protein-protein interactions.  
CC -!- DATABASE: NMR-Database; NMR-Database; NMR-Database;  
NOTE-Entry 11 of June 2001;  
WWW="http://www.expasy.org/spotlight/articles/sptlt011.html".  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M62654; AAA27722.1; -;  
DR EMBL; M62653; AAA27721.1; -;  
DR EMBL; X29345; AAA58246.1; -;  
DR EMBL; X56418; CAA65278.1; -;  
DR PIR; J30692; JQ1514.  
DR PDB; 1B9C; 17-NOV-00.  
DR PDB; 1BFP; 07-JUL-97.  
DR PDB; 1C4F; 14-JUN-00.  
DR PDB; 1EWA; 08-NOV-96.  
DR PDB; 1EMB; 16-JUN-97.  
DR PDB; 1EMC; 20-AUG-97.  
DR PDB; 1EME; 20-AUG-97.  
DR PDB; 1EMF; 20-AUG-97.  
DR PDB; 1EMG; 12-MAY-99.  
DR PDB; 1EMK; 20-AUG-97.  
DR PDB; 1EML; 20-AUG-97.  
DR PDB; 1EMN; 20-AUG-97.  
DR PDB; 1F09; 17-NOV-00.  
DR PDB; 1F0B; 17-NOV-00.  
DR PDB; 1GFL; 11-JAN-97.  
DR PDB; 1GJ; 15-JAN-02.  
DR PDB; 1JUY; 04-JUL-01.  
DR PDB; 1JBZ; 07-JAN-03.  
DR PDB; 1KPS; 28-AUG-02.  
DR PDB; 1KVP; 10-APR-02.

DR PDB; 1KYR; 10-APR-02.  
DR PDB; 1KYS; 10-APR-02.  
DR PDB; 1YPP; 28-OCT-98.  
DR PDB; 2EMD; 20-AUG-97.  
DR PDB; 2EMN; 20-AUG-97.  
DR PDB; 2EMO; 20-AUG-97.  
DR InterPro; IPR000786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFP; 1.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
KW Luminescence; 3D-structure.  
FT CROSSLINK 65 67 5-imidazolone (Ser-Gly).  
FT MOD\_RES 66 66 2,3-DIDEHYDROTYROSINE.  
FT VARIANT 100 100 F -> Y.  
FT VARIANT 108 108 T -> S.  
FT VARIANT 141 141 L -> M.  
FT VARIANT 219 219 V -> I.  
FT CONFLICT 2 2 S -> G (IN REF. 3).  
FT CONFLICT 25 25 H -> Q (IN REF. 2).  
FT CONFLICT 80 80 Q -> R (IN REF. 3).  
FT CONFLICT 157 157 Q -> P (IN REF. 2).  
FT CONFLICT 172 172 E -> K (IN REF. 2).  
FT HELIX 4 8  
FT STRAND 12 22  
FT TURN 23 24  
FT STRAND 25 36  
FT TURN 37 40  
FT STRAND 41 48  
FT TURN 49 50  
FT TURN 57 60  
FT TURN 61 63  
FT HELIX 69 71  
FT STRAND 73 73  
FT HELIX 76 81  
FT HELIX 83 86  
FT TURN 87 90  
FT STRAND 92 100  
FT TURN 101 102  
FT STRAND 105 115  
FT TURN 116 117  
FT STRAND 118 128  
FT TURN 132 133  
FT TURN 135 139  
FT STRAND 141 141  
FT STRAND 148 155  
FT TURN 156 159  
FT STRAND 160 171  
FT TURN 172 173  
FT STRAND 176 187  
FT STRAND 199 208  
FT TURN 211 212  
FT STRAND 217 227  
SQ SEQUENCE 238 AA; 26886 MW; EAS6F21FBFB6E05 CRC64;  
Query Match 97.6%; Score 1242; DB 1; Length 238;  
Best Local Similarity 97.9%; Pred. No. 3.5e-96;  
Matches 233; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQGEKDAATYVKLTSLKICTTGKLPVWPPTL 61  
:|||||  
Db 1 MSKGEELFTGVVPIVLVELDGDVNGHKFVSQGEKDAATYVKLTSLKICTTGKLPVWPPTL 60  
QY 62 VTALSYGVOCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNVKTAEVKEFGDTLV 121  
:|||||  
Db 61 VTTFSGVOCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNVKTAEVKEFGDTLV 120  
QY 122 NRTELKGDIDFKEDGNILGHKLEYNYNHSHVYIMADKQKGIKYNFKIRHNIEDGSVOLAD 181  
:|||||  
Db 121 NRTELKGDIDFKEDGNILGHKLEYNYNHSHVYIMADKQKGIKYNFKIRHNIEDGSVOLAD 180  
QY 182 HYQONTPIGDGPVLLPDNHYLSTQSSALSKDPNKRDMVLLGFVTAAGITLGDYLYK 239  
:|||||

```
Db 181 HQOQNTPIGDPVLLPDNHYLSTQSALSQKDPNEKRDHMLLEFVTAAGITHGMDELYK 238
RESULT 2
SYL_HABIN
ID SYL_HABIN STANDARD; PRT; 861 AA.
AC P43827;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--trNA ligase) (Leurs).
GN LEUS OR H10921.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-P., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32774; AAC22581.1; -;
DR F01; H64102; H64102.
DR TIGR; H10921; -.
DR HAMAP; MF_00049; -; 1.
DR InterPro; IPR002302; Leu-TRNASynt1a.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR009008; ValRS_fleRS_edit.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; P000985; TRNASYNTHLEU.
DR TIGRPFAMS; TIGR00396; leuS bact; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT BINDING 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY)
SQ SEQUENCE 861 AA; 97750 MW; EB93304F6B4C8FB7 CRC64;
Query Match 7.2%; Score 91.5; DB 1; Length 861;
Best Local Similarity 24.1%; Pred. NO. 6.2;
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;
QY 50 TTGKLPVPHPTLTALSYGQCFSRYPDHMKQHDFFKASMPGYGVERIIFPKD----- 103
DB 314 TGDKLPI-WVANFVLMHYGTGAVMAVPAH-DQRDF-----EFAQKYSLPKIQVIAPLA 364
```

```
QY 104 DGNKYTRAEVKEFGDTLVNRIELKGDIFKEDGNILGHKLEYNNYNNHNYIMADK-QKNGI 162
Db 365 DEEIDLTKQAFVEHGKLVNSDFDGKNF--DGAFNG-----IADKLEKLGV 408
QY 163 ---KYNFKIRH-----NIEDSVQLADHYQQTPIGDPVLLPDNHYL- 202
Db 409 GKRQVNYRLRDWGVSRQRYGAPIPMLTLLENGDVVPA-----PMEDLPILPEDVVD 461
QY 203 STQSALSQKDPN 213
Db 462 GVKSPINADPN 472
RESULT 3
SYL_XYLFA
ID SYL_XYLFA STANDARD; PRT; 879 AA.
AC Q9PBG8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--trNA ligase) (Leurs).
GN LEUS OR XF2176.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=9a56;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.B., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.P., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facinelli A.P., Ferreiro A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohne M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hobeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.B., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega P.G., Nunes L.R., Oliveira M.A., Paris A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zaglo M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF004031; AAF84975.1; ALT_INIT.
DR HAMAP; MF_00049; -; 1.
```

```
DR InterPro; IPR002302; Leu-TRNAsyntla.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR009008; ValRS_fIERS_edit.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRFAMs; TIGR00396; leuS_bact; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 45 55 "HIGH" REGION.
FT SITE 637 641 "KMSKS" REGION.
FT BINDING 640 640 ATP (BY SIMILARITY)
SQ SEQUENCE 879 AA; 99796 MW; 9FDCB99202919E CRC64;

Query Match 7.0%; Score 89.5; DB 1; Length 879;
Best Local Similarity 22.2%; Pred. No. 9.4;
Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;

QY 50 TTGKLPVPTLTALSYGQCSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNY-- 107
DB 321 TNEQLVW-VWANEVLMAYGAVMAYVEGHDDQDEF--ANKYGLPIRQVIALKEPKNQDE 377
QY 108 -----KTRAEVKPEGDTLVNRILKGDIDFKEDGNILGHKLEYNYNHNVYI 153
DB 378 STWEPDVRWDYADKTR---EFE--LINSAEFDGLDYQDAFEVLAERFE----- 421
QY 154 MADKQKNG-IKVNFKIRHNIEDSGVOLADHYQNTPI-----GDGPVLLPDN 199
DB 422 ---RQGRQRRVNYRLR----DWGSRQRYWGCPFIVYCTGCAVPEVPEDQLPVLPEN 474
QY 200 -HYLSTQSALSADPNKR 216
DB 475 VAFSGTGSPKTPDEWRK 492

RESULT 4
SY62_DISOM STANDARD; PRT; 439 AA.
ID VIT4_CABEL
AC P18947; Q9BPP3;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Synaptotagmin B (Synaptic vesicle protein O-p65-B).
GN P65-B.
OS Discopyle ommata (Electric ray).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squala; Hyphosqualea; Pristiogadidae; Batoidae;
OC Torpediniformes; Narcinoidei; Narcinidae; Discopyle.
OX NCBI_TaxID=7785;
RN [1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=91273991; PubMed=2054189;
RA Wendland B., Miller K.G., Schilling J., Scheller R.H.;
RT "Differential expression of the p65 gene family.";
RL Neuron 6:993-1007(1991).
CC -1- FUNCTION: May have a regulatory role in the membrane interactions
CC during trafficking of synaptic vesicles at the active zone of the
CC synapse. It binds acidic phospholipids with a specificity that
CC requires the presence of both an acidic head group and a diacyl
CC backbone.
CC -1- SUBUNIT: Homodimer or homotrimer (possible).
CC -1- SUBCELLULAR LOCATION: Synaptic vesicles in neurons.
CC -1- TISSUE SPECIFICITY: Spinal cord, brainstem, midbrain and electric
CC organ.
CC -1- SIMILARITY: Contains 2 C2 domains.
CC -1- SIMILARITY: Belongs to the synaptotagmin family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M64276; AAA49228.1; -.
DR PIR; JH0414; JH0414.
DR HSSP; P21707; 1BYN.
DR InterPro; IPR000008; C2.
DR InterPro; IPR008973; C2_CaLB.
DR InterPro; IPR002149; LRI.
DR Pfam; PF00168; C2; 2.
DR SMART; SM00239; C2; 2.
DR PROSITE; PS00499; C2_DOMAIN_1; 2.
DR PROSITE; PS00004; C2_DOMAIN_2; 2.
KW Transmembrane; Repeat; Synapse; Multigene family; Glycoprotein.
FT DOMAIN 1 74 VESICULAR (POTENTIAL).
FT TRANSMEM 75 101 POTENTIAL.
FT DOMAIN 102 439 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 153 399 PHOSPHOLIPID BINDING (PROBABLE).
FT DOMAIN 173 262 C2 DOMAIN 1.
FT DOMAIN 304 395 C2 DOMAIN 2.
FT CARBOHYD 6 6 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 46 46 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 439 AA; 49278 MW; 2033F05FD8C69F39 CRC64;

Query Match 7.0%; Score 89; DB 1; Length 439;
Best Local Similarity 19.5%; Pred. No. 4.5;
Matches 56; Conservative 46; Mismatches 93; Indels 92; Gaps 12;

QY 16 LVELDGVNHFVSVSGEEDATYKLTLCITCTGKLPV-PWPTLVLTALSFGV----- 69
DB 38 MNPIDTGDNSTAGVPEGKND-VFEKKEKFMNELQKLPPLPWALIAIVSGLLLTLC 96
QY 70 -----QCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKPEG- 117
DB 97 CLICIKKCKCKKKKKKCKGK-----KNDINMK---DVKSGSGNQDDD 138
QY 118 --DTLVNRIELKGDIDFKEDGNI--LGHKLEYNYNH----- 149
DB 139 DAETGLTEGEDKEEAEKEELKGIQFSLDYDFQANALVIGIIQAAELPALDMGTSDPY 198
QY 150 -NVYIMADKQKN-GIKVN-----FKIRHNIEDSGVOL-----DHVQNTPI 189
DB 199 VKVFLPLDPKKKYETKVQKTLNPTNFSFVKVPYQELGGKTLTMNAVYDFDFRSKDCI 258
QY 190 GDGPVLLPD-----NHVLTQSALSADPNKRKDHMLLIGFVTAAG 229
DB 259 GQVTVLMTKVDLQQLLEWRDLSEAEKEPEKLGDICTSLRYVYPTAG 305

RESULT 5
VIT4_CABEL STANDARD; PRT; 1603 AA.
ID VIT4_CABEL
AC P18947; Q9BPP3;
DT 01-NOV-1990 (Rel. 16, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vitellogenin 4 precursor.
GN Vit-4 OR F59D8.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE OF 1-282 FROM N.A.
RA Blumenthal T., Spieth J., Zucker E.;
RL Submitted (NOV-1986) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE OF 1-71 FROM N.A.
RP
```



```

RESULT 7
SYL_SALT
ID SYL_SALT STANDARD; PRT; 860 AA.
AC Q8ZQ26;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR STM0648;
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florio R.K., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856 (2001).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diposphate + L-leucyl-tRNA(Leu).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; AE008725; AL19599.1; -.
CC StyGene; SG????; leus.
CC HAMAP; MF 00049; -.
CC InterPro; IPR002302; Leu-TRNAsyntla.
CC InterPro; IPR002300; tRNA-synt la.
CC InterPro; IPR001412; tRNA-synt I.
CC InterPro; IPR009008; ValRS lIeRS_edit.
CC Pfam; PF00133; tRNA-synt I; 1.
CC PRINTS; PR00985; TRNASYNTHLEU.
CC TIGRFAMS; TIGR00396; leus_bact; 1.
CC PROSITE; PS00178; AA TRNA LIGASE I; 1.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC SITE 42 52 "HIGH" REGION.
CC SITE 619 623 "KMSKS" REGION.
CC BINDING 622 622 ATP (BY SIMILARITY).
CC SEQUENCE 860 AA; 96985 MW; D5003584DFECCAB6 CRC64;

Query Match 6.9%; Score 87.5; DB 1; Length 860;
Best Local Similarity 23.3%; Pred. No. 13;
Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;

QY 50 TTGKLPVPWPTLTALSYGVCFSRYPDHMKQDFKFSAMPEGVQVORTFFKDDGNYKT 109
Db 314 TGEIIPV-WAANFVLMYEGTGAVWVPGH-DQRD-YEFASKYGLTIKPVLAADGSEPD 370
QY 110 RAEVKFGEDTLVNRLEKIDFDKEDGNILGHLEYNYNHNVMADKQNGIKVNFKIR 169
Db 371 SEQALTEKGVLFNSGSEPDGLAFEAFAFNAIADKL-----AEKGVGERKVNYRLR 418
QY 170 H-----NIEDGSQVLADHYQQNTPIGDGPVLAPDNHVL-STQSALSKDP 212
Db 419 DWGVSQRQYWGAPIPWVTLEDGTV-----LPTPEDQLPVLFDVMDGITSPIKADP 471

RESULT 8
CP51_CANGA STANDARD; PRT; 533 AA.
AC P50859; Q02312;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 51 (EC 1.14.13.70) (CYP11) (P450-LIA1) (Sterol 14-
DE alpha-demethylase) (Lanosterol 14-alpha demethylase) (P450-14DM).
GN ERG11 OR CYP51.
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2001-L5;
RX MEDLINE=96161286; PubMed=8593007;
RA Geber A., Hitchcock C.A., Swartz J.E., Pullen F.S., Marsden K.E.,
RA Kwon-Chung K.J., Bennett J.E.;
RA "Deletion of the Candida glabrata ERG3 and ERG11 genes: effect on cell
RT viability, cell growth, sterol composition, and antifungal
RT susceptibility."
RL Antimicrob. Agents Chemother. 39:2708-2717 (1995).
RN [2]
RP SEQUENCE OF 60-473 FROM N.A.
RC STRAIN=ATCC 2001;
RX MEDLINE=95081364; PubMed=7989540;
RA Burgener-Kairuz P., Zuber J.P., Jaunin P., Buchman T.G., Bille J.,
RA Rossier M.;
RA "Rapid detection and identification of Candida albicans and
RT Torulopsis (Candida) glabrata in clinical specimens by
RT species-specific nested PCR amplification of a cytochrome P-450
RT lanosterol-alpha-demethylase (LIA1) gene fragment."
RL J. Clin. Microbiol. 32:1902-1907 (1994).
CC -1- FUNCTION: Catalyzes C14-demethylation of lanosterol which is
CC critical for ergosterol biosynthesis. It transforms lanosterol
CC into 4,4'-dimethyl cholesta-8,14,24-triene-3-beta-ol (By
CC similarity).
CC -1- CATALYTIC ACTIVITY: Obtusifolitol + 3 O(2) + 3 NADPH = 4-alpha-
CC methyl-5-alpha-ergosta-8,14,24(28)-trien-3-beta-ol + formate + 3
CC NADP(+) + 3 H(2)O.
CC -1- PATHWAY: Ergosterol biosynthesis.
CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; L40389; AAB02329.1; -.
CC ENBL; S75389; AAB32679.1; -.
CC InterPro; IPR001128; Cytochrome_P450.
CC Pfam; PF00067; P450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC Electron transport; Oxidoreductase; Monooxygenase; Membrane; Heme;
CC Sterol biosynthesis; NADP.
CC METAL 472 472 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
CC CONFLICT 64 64 I -> M (IN REF. 2).
CC CONFLICT 473 473 I -> T (IN REF. 2).
CC SEQUENCE 533 AA; 61305 MW; A0506C17507E6EF7 CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 533;
Best Local Similarity 21.8%; Pred. No. 9.1;
Matches 45; Conservative 32; Mismatches 80; Indels 49; Gaps 9;

QY 25 GHKFSVS---GEGEGDATYKGLTLKICTGKLPVWPVTLVLTALSYGVCFSRYPDH--M 79

```







CC entities requires a license agreement (See http://www.isb-sib.ch/announcement or send an email to license@isb-sib.ch).

CC  
CC EMBL; AL139076; CAB73096.1; -.  
CC PIR; G81355; G81355.  
DR HAMAP; MF\_01011; ; 1.  
DR InterPro; IPR000051; SAM\_bind.  
DR InterPro; IPR001566; TrmA.  
DR PROSITE; PS01230; TRMA.1; 1.  
DR PROSITE; PS01231; TRMA.2; FALSE NEG.  
KW Transferase; Methyltransferase; tRNA processing; Complete proteome.  
FT S-ADENOSYLMETHIONINE BINDING (BY  
FT SIMILARITY).  
FT ACT\_SITE 315 315 BY SIMILARITY.  
SQ SEQUENCE 357 AA; 42276 MW; CEC5328347CBE497 CRC64;

Query Match 6.8%; Score 86; DB 1; Length 357;  
Best Local Similarity 24.8%; Pred. No. 6.2;  
Matches 30; Conservative 18; Mismatches 39; Indels 34; Gaps 5;

QY 80 KOHDFKFSAMPEGYVQERTIFFKDDGNKYKTRAEVKF--EGDTLV-----NRILKG 128  
Db :|||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
14 EXHSFIKKYKFETDKFLFKASDKHYRTRAEISFYHENDTLFVAFDPDKSKKIIEY 73

QY 129 IDPKED-----GNILGHKLEYNYSNHNVIYIMADKQNGIKVNFIRINIE 173  
Db :|||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
74 LDFADEKICAFMPRLLEYLRQDNKLKEKL-----FGVEFLTTKQE--LSITLLYHKNIIE 125

QY 174 D 174  
Db 126 D 126

RESULT 14

ID DPOL\_HPBHE STANDARD; PRT; 788 AA.

AC F13846;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE P protein [includes: DNA-directed DNA polymerase (EC 2.7.7.7); RNA-directed DNA polymerase (EC 2.7.7.49); Ribonuclease H (EC 3.1.26.4)].  
GN P.  
OS Heron hepatitis b virus.  
OC Viruses; Retroviral viruses; Hepadnaviridae; Avihepadnavirus.  
OX NCBI\_TaxID=28300;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88333160; PubMed=3418788;  
RA Sprengel R., Kalata E.F., Will H.;  
RT Isolation and characterization of a hepatitis B virus endemic in herons";  
RL J. Virol. 62:3832-3839 (1988).  
CC -I- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate + [DNA] (N).  
CC -I- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-phosphomonoester.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).

CC EMBL; M220556; AAA45738.1; -.  
DR PIR; A30082; JDVLHH.  
DR InterPro; IPR001462; DNAPol\_viral\_C.  
DR InterPro; IPR000201; DNAPol\_viral\_N.  
DR InterPro; IPR000477; RVtse.  
DR Pfam; PF00336; DNA\_pol\_viral\_C; 1.  
DR Pfam; PF00242; DNA\_pol\_viral\_N; 1.



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:45:30 ; Search time 30.6667 Seconds  
(without alignments)  
2458.984 Million cell updates/sec

Title: US-09-887-784-64A  
Perfect score: 1273  
Sequence: 1 MVSGEELFTGVVPILVELD.....VLLGFVTRAGITLGMDELYK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL.25:\*
- 1: sp\_archea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phase:\*
  - 10: sp\_plant:\*
  - 11: sp\_prodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertibrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvirus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	1238	97.3	2 Q8GHE2	Q8ghe2 azotobacter
2	1235	97.0	238 5 Q93125	Q93125 aequorea vi
3	1233	96.9	238 2 Q8GHE4	Q8ghe4 azomonas ag
4	1232	96.8	238 2 Q8GHE3	Q8ghe3 azotobacter
5	1200	94.3	238 5 Q17105	Q17105 aequorea vi
6	1185	93.1	238 5 Q17106	Q17106 aequorea vi
7	1080	84.8	238 5 Q8WTC6	Q8wtc6 aequorea ma
8	1076	84.5	238 5 Q8WPC95	Q8wpc95 aequorea ma
9	1072	84.2	238 5 Q8WTC4	Q8wtc4 aequorea ma
10	1070	84.1	238 5 Q8WTD0	Q8wtcd0 aequorea ma
11	1069	84.0	238 5 Q8WTC8	Q8wtc8 aequorea ma
12	1059	84.0	238 5 Q8WTC9	Q8wtc9 aequorea ma
13	1067	83.8	238 5 Q8WTC7	Q8wtc7 aequorea ma
14	1065	83.7	238 5 Q8WTC5	Q8wtc5 aequorea ma
15	256.5	20.1	225 5 Q95UA7	Q95ua7 montastraea
16	256.5	20.1	225 5 Q7Z0W5	Q7z0w5 montastraea

17	251	19.7	225 5	Q963F5	Q963f5 montastraea
18	248.5	19.5	236 5	Q8TFU0	Q8tfu0 dendronephth
19	244	19.2	225 5	Q8I6J8	Q8i6j8 trachyphyl
20	242.5	19.0	225 5	Q7Z0W9	Q7z0w9 montastraea
21	242.5	19.0	266 5	Q9U6Y3	Q9u6y3 clavularia
22	237	18.6	225 5	Q7Z0W4	Q7z0w4 montastraea
23	236	18.5	224 5	Q8MU48	Q8mu48 montastraea
24	232	18.2	225 5	Q8T5F1	Q8t5f1 montastraea
25	216.5	17.0	259 5	Q8MMA2	Q8mma2 agaricia fr
26	216	17.0	239 5	Q8MMA1	Q8mma1 agaricia ag
27	214	16.8	227 5	Q7Z0W6	Q7z0w6 montastraea
28	214	16.8	234 5	Q7Z0W7	Q7z0w7 montastraea
29	209	16.4	234 5	Q8T5F2	Q8t5f2 montastraea
30	209	16.4	234 5	Q8MU47	Q8mu47 montastraea
31	208.5	16.4	229 5	Q9U6Y6	Q9u6y6 anemonia ma
32	207.5	16.3	238 5	Q9BLV9	Q9blv9 renilla mue
33	206	16.2	227 5	Q962P9	Q962p9 montastraea
34	206	16.2	227 5	Q7Z0W8	Q7z0w8 montastraea
35	205.5	16.1	232 5	Q9GP15	Q9gp15 anemonia su
36	204	16.0	221 5	Q95P04	Q95p04 goniotopora t
37	202.5	15.9	222 5	Q7Z168	Q7z168 ceriantthus
38	202.5	15.9	225 5	Q8T6T9	Q8t6t9 radianthus
39	202	15.9	227 5	Q95VT0	Q95vt0 montastraea
40	202	15.9	235 5	Q8T5F0	Q8t5f0 scolymia cu
41	201.5	15.8	225 5	Q9U6Y8	Q9u6y8 discosoma s
42	201.5	15.8	232 5	Q9GZ28	Q9gz28 anemonia s
43	199.5	15.7	214 5	Q86LV7	Q86lv7 meandrina m
44	198.5	15.6	214 5	Q86LV8	Q86lv8 meandrina m
45	197.5	15.5	232 5	Q9U6Y7	Q9u6y7 discosoma s

ALIGNMENTS

RESULT 1

Q8GHE2	ID	Q8GHE2	PRELIMINARY;	PRT;	238 AA.
AC	Q8GHE2;				
DT	01-MAR-2003 (trEMBLrel. 23, Created)				
DT	01-MAR-2003 (trEMBLrel. 23, Last sequence update)				
DT	01-OCT-2003 (trEMBLrel. 25, Last annotation update)				
DE	Green fluorescence protein.				
GS	2289GFP.				
OS	Azotobacter vinelandii.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;				
OC	Pseudomonadaceae; Azotobacter.				
OX	NCBI_TaxID=354;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=DSM2289;				
RA	Koranyi P., Berenyi M., Burg K.;				
RT	"Occurrence of green fluorescence protein in diazotrophic bacteria				
RT	Azomonas and Azotobacter.";				
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF324408; AAN86140.1; -.				
DR	GO; GO:0006091; P:energy pathways; IEA.				
DR	InterPro; IPR009017; GFP like.				
DR	InterPro; IPR000786; Green_fl_protein.				
DR	Fram; PF01353; GFP; 1.				
DR	PRINTS; PR01229; GFLUORESCENT.				
DR	ProDom; PD013756; Green_fl_protein; 1.				
SQ	SEQUENCE 238 AA; 26914 MW; F84840F1F9064018 CRC64;				

Query Match	97.3%;	Score 1238;	DB 2;	Length 238;
Best Local Similarity	97.5%;			
Matches 232;	Conservative	2;	Mismatches 4;	Indels 0; Gaps 0;
Qy	2	VSKGEELFTGVVPILVELDGVNGHKFSVSGEGGDATYGLTLKFICTTGKLPVPMPTL	61	
Db	1	MSKGEELFTGVVPILVELDGVNGHKFSVSGEGGDATYGLTLKFICTTGKLPVPMPTL	60	
Qy	62	VTALSYGVOCFSRYPDPMKHQDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKEGDTLV	121	

Db 61 VTTFSYGVQCFRSYPDHMKRHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKEFGDTLV 120

QY 122 NRIELKGIIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSGVOLAD 181

Db 121 NRIELKGIIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSGVOLAD 180

QY 132 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVLLGFTVTAAGITLGMDELYK 239

Db 131 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVLLGFTVTAAGITLGMDELYK 238

RESULT 2

Q93125 PRELIMINARY; PRT; 238 AA.

AC Q93125; 02, Created

DT 01-FEB-1997 (T-EMBLrel. 02, Last sequence update)

DT 01-FEB-1997 (T-EMBLrel. 02, Last sequence update)

DE 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DE Green fluorescent protein mutant 3.

GN GFP.

OS Aequorea victoria (Jellyfish).

OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;

OC Aequoreidae; Aequorea.

OX NCBI\_TaxID=6100;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96305137; PubMed=8707053;

RA Cormack B.P., Valdivia R.H., Falkow S.;

RT "FACS-optimized mutants of the green fluorescent protein (GFP).";

RL Gene 173:33-38(1996).

RN [2]

RN SEQUENCE FROM N.A.

RA Cormack B.P., Bertram G., Egerton M., Gow N.A.R., Falkow S.,

RA Brown A.J.P.;

RT "Yeast Enhanced Green Fluorescent Protein (yEGFP): a reporter of gene

RT expression in Candida albicans.;"

RL Microbiology 0:0-0(1996).

DR EMBL; U73901; AAB18957.1; -.

DR HSP; P42212; 1BFP.

DR GO; GO:0006091; P:energy pathways; IEA.

DR InterPro; IPR009017; GFP like.

DR InterPro; IPR000786; Green\_fl\_protein.

DR Pfam; PF01353; GFP; 1.

DR PRINTS; PR01229; GFP.

DR ProDom; PD013756; Green fl protein; 1.

DR ProDom; PD013756; Green fl protein; 1.

SQ SEQUENCE 238 AA; 26940 MW; A28622809A9DEA60 CRC64;

Query Match 97.0%; Score 1235; DB 5; Length 238;

Best Local Similarity 97.1%; Pred. No. 2.2e-96;

Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLDGVDVNGHKFSVSGEGDGYKLTGKLTCTTGKLPVWPPTL 61

Db 1 MSKGEELFTGVVPIVLDGVDVNGHKFSVSGEGDGYKLTGKLTCTTGKLPVWPPTL 60

QY 62 VTALSYGVQCFRSYPDHMKRHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKEFGDTLV 121

Db 61 VTTFSYGVQCFRSYPDHMKRHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKEFGDTLV 120

QY 122 NRIELKGIIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSGVOLAD 181

Db 121 NRIELKGIIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVLLGFTVTAAGITLGMDELYK 239

Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVLLGFTVTAAGITLGMDELYK 238

RESULT 3

Q8GHE4 PRELIMINARY; PRT; 238 AA.

AC Q8GHE4; 23, Created

DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DE Green fluorescence protein.

GN 375GFP.

OS Azomonas agilis.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Azomonas.

OX NCBI\_TaxID=116849;

RN [1]

RP SEQUENCE FROM N.A.

RA Koranyi P., Berenyi M., Burg K.;

RT "Occurrence of green fluorescence protein in diazotrophic bacteria

RT Azomonas and Azotobacter.;"

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF324405; AAN86137.1; -.

DR GO; GO:0006091; P:energy pathways; IEA.

DR InterPro; IPR009017; GFP like.

DR InterPro; IPR000786; Green\_fl\_protein.

DR Pfam; PF01353; GFP; 1.

DR PRINTS; PR01229; GFP.

DR ProDom; PD013756; Green fl protein; 1.

DR ProDom; PD013756; Green fl protein; 1.

SQ SEQUENCE 238 AA; 26902 MW; 15FE9B9C5B4F6B89 CRC64;

Query Match 96.9%; Score 1233; DB 2; Length 238;

Best Local Similarity 97.1%; Pred. No. 3.3e-96;

Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLDGVDVNGHKFSVSGEGDGYKLTGKLTCTTGKLPVWPPTL 61

Db 1 MSKGEELFTGVVPIVLDGVDVNGHKFSVSGEGDGYKLTGKLTCTTGKLPVWPPTL 60

QY 62 VTALSYGVQCFRSYPDHMKRHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKEFGDTLV 121

Db 61 VTTFSYGVQCFRSYPDHMKRHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKEFGDTLV 120

QY 122 NRIELKGIIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSGVOLAD 181

Db 121 NRIELKGIIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVLLGFTVTAAGITLGMDELYK 239

Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVLLGFTVTAAGITLGMDELYK 238

```
Best Local Similarity 97.1%; Pred. No. 4e-96;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVELDGVNKHFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPIVLVELDGVNKHFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 60

QY 62 VTALS YGVQCFSRYPDHMKQHDFFKSAMPEGYYQERTIFFKDDGNKTKRAEVKFEGLTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYYQERTIFFKDDGNKTKRAEVKFEGLTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFIRHNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQTSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSQTSALS KDPNEKRDMHVLLEFVTAAGITLGMDELYK 238

RESULT 5
Q17105
ID Q17105 PRELIMINARY; PRT; 238 AA.
AC Q17105
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83959; CAA58789.1; -.
DR PIR; JS0692; J01514.
DR HSSP; P42212; 1GFL.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748E44 CRC64;

Query Match 94.3%; Score 1200; DB 5; Length 238;
Best Local Similarity 93.7%; Pred. No. 2e-93;
Matches 223; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVELDGVNKHFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPIVLVELDGVNKHFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 60

QY 62 VTALS YGVQCFSRYPDHMKQHDFFKSAMPEGYYQERTIFFKDDGNKTKRAEVKFEGLTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGYYQERTIFFKDDGNKTKRAEVKFEGLTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFIRHNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQTSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSQTSALS KDPNEKRDMHVLLEFVTAAGITLGMDELYK 238

RESULT 6
Q17106
ID Q17106 PRELIMINARY; PRT; 238 AA.
AC Q17106
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83959; CAA58789.1; -.
DR PIR; JS0692; J01514.
DR HSSP; P42212; 1GFL.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748E44 CRC64;

Query Match 93.1%; Score 1185; DB 5; Length 238;
Best Local Similarity 92.9%; Pred. No. 3.8e-92;
Matches 221; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVELDGVNKHFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPIVLVELDGVNKHFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 60

QY 62 VTALS YGVQCFSRYPDHMKQHDFFKSAMPEGYYQERTIFFKDDGNKTKRAEVKFEGLTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGYYQERTIFFKDDGNKTKRAEVKFEGLTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFIRHNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQTSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSQTSALS KDPNEKRDMHVLLEFVTAAGITLGMDELYK 238

RESULT 7
Q8WTC6
ID Q8WTC6 PRELIMINARY; PRT; 238 AA.
AC Q8WTC6
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPxm19uv;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435431; AAL33916.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP_like.
DR ProDom; PD013756; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
```

```
DR PRINTS: PR01229; GFP, FLUORESCENT.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27015 MW; 688FD75E88926903 CRC64;

Query Match      84.8%; Score 1080; DB 5; Length 238;
Best Local Similarity 81.9%; Pred. No. 2.8e-83;
Matches 195; Conservative 20; Mismatches 23; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSFGEGDGYKGLTKLFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGIVPVLIELDGDVHGKFSVRGEGDADYKGLFICTTGKLPVWPPTL 60

QY 52 VTALSVCVCFSPYDPHMKQHDFFKSAFPEGYQERTIPFKDGNKYKTRAEVKFEGDTLV 121
DB 51 VTTLSVGIQCFARYPEHMKNDFFKSAFPEGYIQERTIPFDGCKYKTRGEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
DB 121 NRIELKGMDFKEDGNILGHKLEYNHNHYIMPDKANNGLKVNFKIRHNIEGGVOLAD 180

QY 132 HYQONTPIGDGPVLLPDNHYLSQTSALSADPNKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 131 HYQTNVPLGDGPVLLPINHYLSQTALSKDRNETRDHMLVLEFFSACGHTGMDLYK 238

RESULT 8
Q8WP95 PRELIMINARY; PRT; 238 AA.
AC Q8WP95;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFPXM.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL33918.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP, 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match      84.2%; Score 1072; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 1.3e-82;
Matches 196; Conservative 17; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSFGEGDGYKGLTKLFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGDVHGKFSVRGEGDADYKGLFICTTGKLPVWPPTL 60

QY 62 VTALSVCVCFSPYDPHMKQHDFFKSAFPEGYQERTIPFKDGNKYKTRAEVKFEGDTLV 121
DB 61 VTTLSVGIQCFARYPEHMKNDFFKSAFPEGYIQERTIPFDGCKYKTRGEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
DB 121 NRIELKGMDFKEDGNILGHKLEYNHNHYIMPDKANNGLKVNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQTSALSADPNKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 181 HYQTNVPLGDGPVLLPINHYLSQTALSKDRNETRDHMLVLEFFSACGHTGMDLYK 238

RESULT 10
Q8WTD0 PRELIMINARY; PRT; 238 AA.
AC Q8WTD0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL33918.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP, 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match      84.5%; Score 1076; DB 5; Length 238;
Best Local Similarity 81.5%; Pred. No. 6.1e-83;
Matches 194; Conservative 20; Mismatches 24; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSFGEGDGYKGLTKLFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGIVPVLIELDGDVHGKFSVRGEGDADYKGLFICTTGKLPVWPPTL 60

QY 52 VTALSVCVCFSPYDPHMKQHDFFKSAFPEGYQERTIPFKDGNKYKTRAEVKFEGDTLV 121
DB 51 VTTLSVGIQCFARYPEHMKNDFFKSAFPEGYIQERTIPFDGCKYKTRGEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
DB 121 NRIELKGMDFKEDGNILGHKLEYNHNHYIMPDKANNGLKVNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQTSALSADPNKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 181 HYQTNVPLGDGPVLLPINHYLSQTALSKDRNETRDHMLVLEFFSACGHTGMDLYK 238
```



```

Db      1  MSKGEELFTGVVPIILVELDGDVHGKFSVRKEGEGDADYGLKLEKICTTGTGKLPVWPPTL 60
Qy      62  VTALSYGVOCFRYPDHMKQHDFFPKSAMPEGYVOERTIEFKDDGNKYKTRAEVKFEGDTLV 121
Db      61  VTTLGYGILCFARYPEHMKQNDFFPKSAMPEGYIOERTIEFFQDDGKYKTRGEVKFEGDTLV 120
Qy      122  NRTELKGIQKEDGNLTGHKLEYNYNHNVYIMADKQNGIKVNFIRHNIEDGSVQLAD 181
Db      121  NRTELKGMDFKEDGNLTGHKLEYNFNHNVYIMPDKANGLKVNFKIRHNIEGGVQLAD 180
Qy      182  HYQONTPIGDGPVLLPDNHYLSTQSALSQKPNKRDHMYLLGVPTAAGITLGMDELYK 239
Db      181  HYQTNVPLGDGPVLIIPINHLYSYQTAKSKORNETRDHMYVLEPFSACGHTGMDLYK 238

RESULT 15
ID Q95UA7 PRELIMINARY; PRT; 225 AA.
QY Q95UA7;
AC 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cyan fluorescent protein (Fragment).
OS Montastraea cavernosa (Great star coral).
OC Eukaryota; Metazoa; Chordata; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastraea.
OX NCBI_TaxID=63558;
RN [1]
RP SEQUENCE FROM N.A.
RA Falkowski P.G., Sun Y.
RL "Montastraea cavernosa fluorescent protein.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY056460; AAL17905.1; -.
DR GO: GO:0006091; P:energy pathways; IEA.
DR InterPro: IPR009017; GFP_like.
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRINTS: PR01229; GFP_LUOESCENT.
DR ProDom: PD013756; Green_fl_protein; 1.
FT NON TER 225
SQ SEQUENCE 225 AA; 25775 MW; 52DE2F716D083524 CRC64;

Query Match 20.1%; Score 256.5; DB 5; Length 225;
Best Local Similarity 31.6%; Pred. No. 9.9e-14;
Matches 65; Conservative 43; Mismatches 81; Indels 17; Gaps 7;

Qy      12  VVPIILVELDGDVNGHKFSVSGEGDGYATYGLKTLKF-ICTTGTGKLPVWPPTLIVTALSQVQ 70-
Db      7  VMKIKLRMDGIVNGHKFNITGEGKGFPEGTHIILKVKGGPLPFAYDILITAFQYGNR 66
Qy      71  CFGRYPDHMKQHDFFPKSAMPEGYVOERTIEFKDDGNKYKTRAEVKFEGDTLVNRTELKGI 130
Db      67  VFTKYPKDIP-DYFKQSPFEGYSWERSMTFDEQGVCTVTSDIKLEGDCFPFYERFYGVN 124
Qy      131  FKEDGNLTGHK-LEYNYNHNVYIMADKQNGIKVNFIRHNIEDGSVQLADHYQQNTPI 189
Db      125  FPSSGPPYMQKTLKWPSTENMYV-----RDVGLGVDSRTLLEGD-----KHHRCNFRS 175
Qy      190  GDGP---VLLPDNHYLSTQ-SALSQK 211
Db      176  TYGAKKGVLPEYHFVDHRIELSHD 201

Search completed: June 21, 2004, 16:00:08
Job time : 30.7778 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:41:49 ; Search time 47.1111 Seconds  
(without alignments)  
1433.395 Million cell updates/sec

Title: US-09-887-784-64G  
Perfect score: 1275  
Sequence: 1 MYSKGEELTGVVPLVELD.....VLLGFVTAAGITLGMDELK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES										
Result No.	Score	Match	Length	DB ID	Query %	Description				
1	1267	99.4	239	5	AAE17518	Enhanced				
2	1267	99.4	363	6	ABR40352	Human ami				
3	1267	99.4	893	4	AAG65781	Amino aci				
4	1267	99.4	1132	4	AAG65782	Amino aci				
5	1259	98.7	239	5	AAE17517	Enhanced				
6	1256	98.5	239	3	ABE22882	Enhanced				
7	1256	98.5	239	3	AAE54349	Amino aci				
8	1256	98.5	239	3	AAE75584	EGFP sign				
9	1256	98.5	239	4	ABE50804	Jellyfish				
10	1256	98.5	239	4	ABE85900	A. victor				
11	1256	98.5	239	4	ABE31171	Amino aci				
12	1256	98.5	239	5	AAE66198	A. victor				
13	1256	98.5	239	5	ABG94444	Protease				
14	1256	98.5	239	5	AAE14599	Aequorea				
15	1256	98.5	239	6	AAE34958	Aequorea				
16	1256	98.5	239	6	AAE79829	Green flu				
17	1256	98.5	239	6	ABR83616	Green flu				
18	1256	98.5	239	6	ADA38074	Aequorea				
19	1256	98.5	239	7	ABU63204	Aequorea				
20	1256	98.5	239	7	ADC18358	EGFP (enh				
21	1256	98.5	239	7	ABW00914	Aequorea				
22	1256	98.5	239	7	ADE28570	Enhanced				
23	1256	98.5	246	7	ABM79011	Enhanced				
24	1256	98.5	248	5	AAE68319	Jellyfish				
25	1256	98.5	259	5	AAU99804	Biomembra				

26	1256	98.5	265	2	AAW97451	Wild-type
27	1256	98.5	268	5	AAU99803	Biomembra
28	1256	98.5	270	5	AAU99802	Biomembra
29	1256	98.5	272	5	AAU99800	Biomembra
30	1256	98.5	273	5	AAU99801	Biomembra
31	1256	98.5	280	5	AAU99807	Biomembra
32	1256	98.5	281	3	AAU50142	Green flu
33	1256	98.5	281	3	AAE24252	EGFP-MODC
34	1256	98.5	281	5	AAU10888	EGFP-MODC
35	1256	98.5	286	7	ADE28562	EGFP/ hum
36	1256	98.5	289	7	ADE28564	EGFP/ hum
37	1256	98.5	290	7	ADZ28568	EGFP/ hum
38	1256	98.5	290	7	ADZ28566	EGFP/ hum
39	1256	98.5	294	3	AAE22860	EGFP/ hum
40	1256	98.5	294	3	AAE22860	EGFP/ hum
41	1256	98.5	294	5	ABG94422	Recombina
42	1256	98.5	308	2	AAE42181	EGFP/DRM
43	1256	98.5	320	6	ABR83620	HUB1-GFP
44	1256	98.5	323	3	AAE54359	GFP mutan
45	1256	98.5	323	6	ABR83621	RUB1-GFP

ALIGNMENTS

RESULT 1	
AAE17518	
ID	AAE17518 standard; protein; 239 AA.
XX	
AC	AAE17518;
XX	
DT	22-APR-2002 (first entry)
XX	
DE	Enhanced F64L-E222G jellyfish green fluorescent protein mutant.
XX	
XX	Jellyfish; green fluorescent protein; GFP; protein redistribution;
KW	cellular function; genetic reporter; mutant; Stoke's shift; mutoin.
XX	
OS	Aequorea victoria.
OS	Synthetic.
XX	
FT	Key
FT	Misc-difference 65
FT	/note= "Wild type Phe substituted with Leu; This corresponds to position 64 in the wild type protein"
FT	Misc-difference 223
FT	/note= "Wild type Glu substituted with Gly; This corresponds to position 222 in the wild type protein"
FT	
XX	
XX	WO200198338-A2.
PN	
XX	
PD	27-DEC-2001.
XX	
PF	18-JUN-2001; 2001WO-EP006848.
XX	
PR	19-JUN-2000; 2000DK-00000953.
PR	20-JUN-2000; 2000US-0212681P.
PR	10-MAY-2001; 2001DK-00000739.
PR	10-MAY-2001; 2001US-0290170P.
XX	
XX	(BIOI-) BIOIMAGE AS.
PA	
PI	Bjorn SP, Pagliaro L, Thastrup O;
XX	
DR	WPI; 2002-098224/13.
DR	N-PSDB; AAD28163.
XX	
PT	Novel fluorescent protein in in vitro assay for measuring protein kinase activity or dephosphorylation activity, or for measuring protein redistribution, has a green fluorescent protein with F64L and E222G mutation.
PT	
XX	
PS	Claim 9; Page 37; 41pp; English.

XX The invention relates to a fluorescent protein derived from green  
CC fluorescent protein (GFP) or its analogue. The GFP containing mutations  
CC at F64L and E222G has a bigger compared to other GFP's making it very  
CC suitable for high throughput screening due to better resolution. The  
CC fluorescent protein is useful in *in vitro* assays for measuring protein  
CC kinase activity or dephosphorylation activity, or for measuring protein  
CC redistribution. The fluorescent protein is useful in studying cellular  
CC functions in living cells; as protein tags in transgenic animals, living  
CC and fixed cells; organelle tags, secretion marker and genetic reporter.  
CC The fluorescent protein is also useful as a cell or organelle integrity  
CC marker, a marker for changes in cell morphology, as transfection marker,  
CC and as a marker to be used in combination with fluorescence activated  
CC cell sorting (FACS). The novel proteins can also be used as reporters to  
CC monitor live or dead biomass of organisms, such as fungi. The fluorescent  
CC protein is also useful as markers in transcriptional and translational  
CC fusions for performing transposon vector mutagenesis and as a reporter  
CC for bacterial detection. Transposons encoding the fluorescent protein are  
CC useful for screening promoters and for tagging plasmids and chromosomes.  
CC The fluorescent protein engineered into the genome of a phage is useful  
CC for designing diagnostic tool. The present sequence is a DNA encoding  
CC enhanced F64L-E222G jellyfish green fluorescent protein (GFP) mutant  
XX  
SQ Sequence 239 AA;

Query Match 99.4%; Score 1267; DB 5; Length 239;  
Best Local Similarity 99.8%; Pred. No. 1.3e-122;  
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60

QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 2  
ABR40352 standard; protein; 363 AA.  
XX ABR40352;  
AC ABR40352;  
XX  
DT 08-JUL-2003 (first entry)  
DE Human amino acid sequence SEQ ID NO: 6.  
XX Human; heterologous conjugate; intracellular protein.  
XX Homo sapiens.  
OS Aequoria victoria.  
XX WO2003029827-A2.  
PN 10-APR-2003.  
XX  
XX 01-OCT-2002; 2002WO-DK000651.  
XX  
XX 01-OCT-2001; 2001DK-00001433.  
PR 11-OCT-2001; 2001US-0328896P.  
XX  
XX (BIOL-) BIOIMAGE AS.  
PA Terry BR, Nielsen SJ;  
XX  
PI

XX WPI: 2003-430211/40.  
DR N-PSDB; ACC72604.  
XX  
PT Novel cell for identifying modulators of protein interaction, contains a  
PT first conjugate comprising anchor protein, second conjugate having type B  
PT interactor protein and third conjugate with detectable group.  
XX  
PS Disclosure; Page 112-113; 118pp; English.  
XX  
XX The invention relates to a novel cell, comprising three heterologous  
CC conjugates (HC), a first HC (HC1) comprising an anchor protein that  
CC specifically binds to an internal structure within the cell conjugated to  
CC an interactor protein (IP) of type A, a second HC (HC2) comprising IP of  
CC type B conjugated to a first protein of interest, and a third HC (HC3)  
CC comprising a second protein of interest conjugated to detectable group.  
CC The cell is useful for detecting if a compound disrupts or induces the  
CC interaction between two intracellular proteins. The cell is also useful  
CC for screening compounds that modulate the interaction between two  
CC intracellular proteins. The present sequence is used in the  
CC exemplification of the invention  
XX  
SQ Sequence 363 AA;

Query Match 99.4%; Score 1267; DB 6; Length 363;  
Best Local Similarity 99.6%; Pred. No. 2.4e-122;  
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60

QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 3  
AAG65781  
ID AAG65781 standard; protein; 893 AA.  
XX AAG65781;  
AC AAG65781;  
XX  
DT 07-JAN-2002 (first entry)  
DE Amino acid sequence of HSPDE4A1-E222G fusion protein.  
XX  
XX PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;  
XX autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;  
XX fusion protein.  
XX Homo sapiens.  
OS Aequorea victoria.  
XX WO200179526-A2.  
PN 25-OCT-2001.  
XX  
XX 11-APR-2001; 2001WO-DK000264.  
XX  
XX 17-APR-2000; 2000DK-00000651.  
PR 29-MAY-2000; 2000DK-00000849.  
XX  
XX (BIOI-) BIOIMAGE AS.  
PA  
XX

```

PI Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
PI Praestegaard M;
XX WPI; 2001-611727/70.
DR N-PSDB; AAI66852.
DR XX
XX Determining if a compound is a dislocator of PDE4 for identifying
PT compounds for treating CNS and inflammatory disease comprises identifying
PT compounds which remove PDE4 spots.
XX
XX Example 1; Page 156-160; 160pp; English.
XX
XX The invention relates to determining, if a compound, is a dislocator of
XX PDE4. The method comprises testing if the compound removes PDE4 spots,
XX which may optionally be induced by a Rolipram-like reference compound,
XX and testing if it inhibits the catalytic activity of the PDE4, where the
XX compound is a dislocator of PDE4, if it removes PDE spots and if it does
XX not inhibit the catalytic activity of PDE4. The method is useful for
XX identifying compounds useful for the treatment of diseases of the central
XX nervous system such as depression and for the treatment of inflammatory
XX disease such as joint inflammation, Crohn's disease, inflammatory bowel
XX disease, respiratory diseases, chronic obstructive pulmonary disease
XX (COPD), including asthma, chronic bronchitis, pulmonary emphysema,
XX endotoxic shock, toxic shock syndrome, systemic lupus erythematosus,
XX psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
XX infection. The use of a reagent that can mimic or reverse the effect of
XX the compound with affinity for the catalytic site on intracellular
XX distribution of the PDE for the preparation of a medicament. The present
XX sequence represents the amino acid sequence of a HSPDE4A1-E222G fusion
XX protein
XX
XX Sequence 893 AA;
XX
XX Query Match 99.4%; Score 1267; DB 4; Length 893;
XX Best Local Similarity 99.6%; Pred. No. 9.1e-122;
XX Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 MVSKEBELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLPFICTTGKLPVPWPT 60
Db 655 MVSKEBELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLPFICTTGKLPVPWPT 714
XX
QY 61 LVTGLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 715 LVTGLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 774
XX
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 775 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 834
XX
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 835 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 893
XX
RESULT 4
AAG65782
ID AAG65782 standard; protein; 1132 AA.
XX
AC AAG65782;
XX
DT 07-JAN-2002 (first entry)
XX
XX Amino acid sequence of HSPDE4A4-E222G fusion protein.
XX
XX PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;
KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;
KW fusion protein.
XX
XX Homo sapiens.
OS Aequorea victoria.
XX
XX WO200179526-A2.
PN
XX

```

---

```

PD 25-OCT-2001.
XX
XX 11-APR-2001; 2001WO-DK000264.
XX
XX 17-APR-2000; 2000DK-00000651.
XX
XX 29-MAY-2000; 2000DK-00000849.
XX
XX (BIOI-) BIOIMAGE AS.
XX
XX Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
XX Praestegaard M;
XX
XX WPI; 2001-611727/70.
DR N-PSDB; AAI66853.
XX
XX Determining if a compound is a dislocator of PDE4 for identifying
PT compounds for treating CNS and inflammatory disease comprises identifying
PT compounds which remove PDE4 spots.
XX
XX Example 1; Page 162-167; 160pp; English.
XX
XX The invention relates to determining, if a compound, is a dislocator of
XX PDE4. The method comprises testing if the compound removes PDE4 spots,
XX which may optionally be induced by a Rolipram-like reference compound,
XX and testing if it inhibits the catalytic activity of the PDE4, where the
XX compound is a dislocator of PDE4, if it removes PDE spots and if it does
XX not inhibit the catalytic activity of PDE4. The method is useful for
XX identifying compounds useful for the treatment of diseases of the central
XX nervous system such as depression and for the treatment of inflammatory
XX disease such as joint inflammation, Crohn's disease, inflammatory bowel
XX disease, respiratory diseases, chronic obstructive pulmonary disease
XX (COPD), including asthma, chronic bronchitis, pulmonary emphysema,
XX endotoxic shock, toxic shock syndrome, systemic lupus erythematosus,
XX psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
XX infection. The use of a reagent that can mimic or reverse the effect of
XX the compound with affinity for the catalytic site on intracellular
XX distribution of the PDE for the preparation of a medicament. The present
XX sequence represents the amino acid sequence of a HSPDE4A4-E222G fusion
XX protein
XX
XX Sequence 1132 AA;
XX
XX Query Match 99.4%; Score 1267; DB 4; Length 1132;
XX Best Local Similarity 99.6%; Pred. No. 1.3e-121;
XX Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 MVSKEBELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLPFICTTGKLPVPWPT 60
Db 894 MVSKEBELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLPFICTTGKLPVPWPT 953
XX
QY 61 LVTGLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 954 LVTGLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 1013
XX
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 1014 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 1073
XX
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 1074 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 1132
XX
RESULT 5
AAEI7517
ID AAEI7517 standard; protein; 239 AA.
XX
XX AAEI7517;
XX
XX 22-APR-2002 (first entry)
XX
XX Enhanced F64L jellyfish green fluorescent protein mutant.
XX

```

KW Jellyfish; green fluorescent protein; GFP; protein redistribution;  
 KW cellular function; genetic reporter; mutant; Stoke's shift; mutain.  
 XX Aequorea victoria.  
 OS Synthetic.  
 XX FH Key Location/Qualifiers  
 FT Misc-difference 65 /note= "Wild type Phe substituted with Leu; This  
 FT corresponds to position 64 in the wild type protein"  
 PN WO200198338-A2.  
 XX PD 27-DEC-2001.  
 XX PF 19-JUN-2001; 2001WO-EP006848.  
 XX PR 19-JUN-2000; 2000DK-00000953.  
 PR 20-JUN-2000; 2000US-0212681P.  
 PR 10-MAY-2001; 2001DK-00000739.  
 PR 10-MAY-2001; 2001US-0290170P.  
 XX (BIOI-) BIOIMAGE AS.  
 PA Bjorn SP, Pagliaro L, Thastrup O;  
 PI WPI; 2002-098224/13.  
 XX DR N-PSDB; AAD28162.  
 XX Novel fluorescent protein in in vitro assay for measuring protein kinase  
 PT activity or dephosphorylation activity, or for measuring protein  
 PT redistribution, has a green fluorescent protein with F64L and E222G  
 PT mutation.  
 XX Example 1; Page 35; 41pp; English.  
 XX The invention relates to a fluorescent protein derived from green  
 CC fluorescent protein (GFP) or its analogue. The GFP containing mutations  
 CC at F64L and E222G has a bigger compared to other GFP's making it very  
 CC suitable for high throughput screening due to better resolution. The  
 CC fluorescent protein is useful in invitro assays for measuring protein  
 CC kinase activity or dephosphorylation activity, or for measuring protein  
 CC redistribution. The fluorescent protein is useful in studying cellular  
 CC functions in living cells; as protein tags in transgenic animals, living  
 CC and fixed cells; organelle tags, secretion marker and genetic reporter.  
 CC The fluorescent protein is also useful as a cell or organelle integrity  
 CC marker, a marker for changes in cell morphology, as transfection marker,  
 CC and as a marker to be used in combination with fluorescence activated  
 CC cell sorting (FACS). The novel proteins can also be used as reporters to  
 CC monitor live or dead biomass of organisms, such as fungi. The fluorescent  
 CC protein is also useful as markers in transcriptional and translational  
 CC fusions for performing transposon vector mutagenesis and as a reporter  
 CC for bacterial detection. Transposons encoding the fluorescent protein are  
 CC useful for screening promoters and for tagging plasmids and chromosomes.  
 CC The fluorescent protein engineered into the genome of a phage is useful  
 CC for designing diagnostic tool. The present sequence is enhanced F64L  
 CC jellyfish green fluorescent protein (GFP) mutant  
 XX Sequence 239 AA;  
 SQ  
 Query Match 98.7%; Score 1259; DB 5; Length 239;  
 Best Local Similarity 99.2%; Pred. No. 8.7e-122;  
 Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MVSGBELFTGVVPIILVELDGVNKHFSVSGEGDATYKLTLPICITTKGLPVPWPT 60  
 DB 1 MVSGBELFTGVVPIILVELDGVNKHFSVSGEGDATYKLTLPICITTKGLPVPWPT 60  
 QY 61 LVTLGSLGVQCFGRYPDPMKHQDFFKSAPEGVVQERTIFFKDDGNVYKTRAEVKEGDTL 120  
 DB 61 LVTTLSGVQCFGRYPDPMKHQDFFKSAPEGVVQERTIFFKDDGNVYKTRAEVKEGDTL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNVNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

Db 121 VNRIELKGIDFKEDGNILGHKLEYNVNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLCFVTAAGITLGMDELYK 239  
 DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLCFVTAAGITLGMDELYK 239

RESULT 6  
 AAB22882  
 ID AAB22882 standard; protein; 239 AA.  
 XX AAB22882;  
 XX DT 10-JAN-2001 (first entry)  
 XX Enhanced green fluorescent protein (EGFP), SEQ ID NO:46.  
 DE Bioreactor protein; fusion protein; recognition site;  
 KW cellular targeting sequence; cellular localisation; fluorescent protein;  
 KW protease activity detection; toxin detection; cellular stress detection;  
 KW drug discovery; cell based screening.  
 XX Aequorea victoria.  
 OS Synthetic.  
 XX WO200050872-A2.  
 XX 31-AUG-2000.  
 XX 25-FEB-2000; 2000WO-US004794.  
 XX 26-FEB-1999; 99US-0122152P.  
 PR 08-MAR-1999; 99US-0123399P.  
 PR 12-JUL-1999; 99US-00352171.  
 XX (CELL-) CELLOMICS INC.  
 PA Giuliano KA, Kapur R;  
 PI WPI; 2000-594086/56.  
 XX N-PSDB; AAA93373.  
 DR Automated cell-based characterization of toxin by contacting cells  
 PT containing luminescent reporter molecules with test substance and  
 PT analyzing optically.  
 XX Example 11; Fig 29A; 336pp; English.  
 XX The invention relates to systems, methods and reagents for cell-based  
 CC screening or detection of compounds which affect particular biological  
 CC functions. The methods of the invention utilise fluorescent bioreactor  
 CC molecules which, when acted on by a compound of interest, cause an  
 CC alteration in the cellular distribution of at least the fluorescent  
 CC moiety. In one embodiment, the biosensors comprise heat shock proteins  
 CC (HSPs) fused to a fluorescent protein (e.g., jellyfish green fluorescent  
 CC protein (GFP), or derivatives thereof). Such biosensors are located in  
 CC the cytoplasm, but on stress activation translocate to the nucleus. In  
 CC another embodiment bioreactor proteins can be used to detect protease  
 CC activity. Such protease bioreactor fusion proteins comprise one or more  
 CC fluorescent proteins; a recognition signal which is cleaved by the  
 CC protease; and at least one cellular localisation signal. The latter two  
 CC components may be components of a single protein which is acted upon by  
 CC the protease, or may be from heterologous sources. Due to the  
 CC localisation signal, the bioreactor protein is localised to a particular  
 CC region of the cell. Once acted on by the protease of interest, the  
 CC fluorescent protein is cleaved from the localisation sequence, and is  
 CC free to migrate to other locations within the cell. The presence of a  
 CC second localisation signal attached to the fluorescent protein enables  
 CC the fluorescent protein to be directed to a different cellular  
 CC compartment after cleavage of the protease recognition sequence. The  
 CC change in distribution of the fluorescent protein can be detected using  
 CC imaging methods with a high degree of spatial resolution. The methods and





PD 20-SEP-2001.  
 XX  
 PF 14-MAR-2001; 2001WO-US008071.  
 XX  
 PR 15-MAR-2000; 2000US-0189698P.  
 XX  
 XX (MERI ) MERCK & CO INC.  
 PA  
 XX Marsh DJ;  
 XX  
 XX WPI; 2001-565791/63.  
 DR N-PSDB; AAH47304.  
 XX  
 XX Fusion proteins comprising melanin concentrating hormone receptor  
 PT peptides and fluorescent proteins, useful for identifying appetite  
 PT stimulants.  
 XX  
 XX Claim 2; Page 14; 71pp; English.  
 XX  
 XX The invention provides melanin concentrating hormone (MCH) receptor  
 CC (MCHR) chimeric and fusion proteins. The MCHR chimeric proteins comprise  
 CC MCHR polypeptide regions from different species. The MCHR fusion protein  
 CC comprise MCHR polypeptide region and a fluorescent polypeptide region  
 CC joined directly, or via a linker, to the carboxy side of the MCHR  
 CC polypeptide region. The MCHR fusion proteins can be expressed by standard  
 CC recombinant methodology. MCH action promotes feeding (orexigenic) and up  
 CC regulation of MCH activity stimulates food intake. The present sequence  
 CC represents a A. victoria green fluorescent protein (GFP) and a linker  
 CC sequence  
 XX  
 XX Sequence 239 AA;  
 SQ  
 Query Match 98.5%; Score 1256; DB 4; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 1.8e-121;  
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKLTLFICTTGTGLPVPWPT 60  
 Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKLTLFICTTGTGLPVPWPT 60  
 QY 61 LVTGLSYGQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 Db 61 LVTTLTYGQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
 Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
 QY 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239  
 Db 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239  
 RESULT 11  
 AAB31171  
 ID AAB31171 standard; protein; 239 AA.  
 XX  
 AC AAB31171;  
 XX  
 XX 02-APR-2001 (first entry)  
 DT  
 XX Amino acid sequence of a green fluorescent protein (GFP).  
 DE  
 XX Growth rate; death rate; reporter gene; luminescent protein;  
 KW fluorescent product; luciferase; green fluorescent protein; GFP.  
 KW  
 XX Aequorea victoria.  
 OS  
 XX WO200075367-A1.  
 PN  
 XX 14-DEC-2000.  
 PD  
 XX 07-JUN-2000; 2000WO-FI000507.  
 PF

XX  
 PR 07-JUN-1999; 99FI-00001296.  
 XX  
 PA (LILI/) LILIUS E.  
 PA (VIRT/) VIRTA M.  
 XX  
 XX Lilius E, Virta M;  
 PI WPI; 2001-061737/07.  
 XX N-PSDB; AAC86954.  
 DR  
 DR Assessing growth and death rates of a micro-organism in a desired  
 XX environment, by introducing 2 reporter genes encoding luminescent and  
 PT fluorescent products and detecting luminescent fluorescence.  
 PT  
 XX Disclosure; Page 27; 32pp; English.  
 PS  
 XX The specification describes a method for assessing the growth rate and  
 CC death rate of a micro-organism within a predetermined time period in a  
 CC desired environment. The method comprises introducing at least two  
 CC reporter genes encoding luminescent and/or fluorescent products into the  
 CC micro-organisms, incubating the micro-organism within the desired  
 CC environment, and detecting luminescence and/or fluorescence after a  
 CC predetermined time period. Use of two different markers within a micro-  
 CC organism enables the differentiation between growth and death rates. The  
 CC method is used to assess the growth rate and death rate of a micro-  
 CC organism within a predetermined time period in a desired environment. The  
 CC present sequence represents a green fluorescent protein (GFP), and is  
 CC encoded by a plasmid which encodes luminescent and fluorescent proteins,  
 CC and is used in the method of the invention  
 CC  
 XX  
 XX Sequence 239 AA;  
 SQ  
 Query Match 98.5%; Score 1256; DB 4; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 1.8e-121;  
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKLTLFICTTGTGLPVPWPT 60  
 Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKLTLFICTTGTGLPVPWPT 60  
 QY 61 LVTGLSYGQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 Db 61 LVTTLTYGQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
 Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
 QY 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239  
 Db 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239  
 RESULT 12  
 AAG66198  
 ID AAG66198 standard; protein; 239 AA.  
 XX  
 AC AAG66198;  
 XX  
 XX 17-JUN-2002 (first entry)  
 DT  
 XX A. victoria green fluorescent protein (EGFP).  
 DE  
 XX Cyan-green fluorescent protein; fluorescence; recombinant; GFP;  
 KW green fluorescent protein; EGFP.  
 KW  
 XX Aequorea victoria.  
 OS  
 XX JP2002045189-A.  
 PN  
 XX 12-FEB-2002.  
 PD  
 XX

```
PF 04-AUG-2000; 2000JP-00237165.
XX
PR 04-AUG-2000; 2000JP-00237165.
XX
XX (RIKA ) RIKAGAKU KENKYUSHO.
XX
XX WPI; 2002-299190/34.
DR N-PSDB; ABL40628.
XX
XX A gene encoding cyan-green fluorescent protein.
PT
XX
XX Example; Page 14; 20pp; Japanese.
XX
CC The invention relates to a gene encoding proteins having cyan-green
CC fluorescence characteristic and having a function of showing stable
CC fluorescence characteristic in acid region. A method for the preparation
CC of a cyan-green fluorescent protein is provided which involves a
CC transformant transformed by a recombinant vector comprising the gene,
CC where the transformant is cultured and the protein is collected from the
CC culture. The present sequence represents the A. victoria green
CC fluorescent protein (EGFP)
XX
XX Sequence 239 AA;
SQ
    Query Match      98.5%; Score 1256; DB 5; Length 239;
    Best Local Similarity 98.7%; Pred. No. 1.8e-121;
    Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
QY 61 LVTGLSYGVCFSRYPDHMKQHDFFKSAPEGVYQVETIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQVETIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
RESULT 13
ABG94444
ID ABG94444 standard; protein; 239 AA.
XX
AC ABG94444;
XX
XX 27-NOV-2002 (first entry)
XX
XX Protease biosensor signal sequence #6.
DE
XX
XX Detection; classification; identification; toxin detection; protease;
XX ADP-ribosylating toxin; cytotoxic phospholipase; exfoliative toxin;
XX toxic threat agent.
XX
XX Synthetic.
XX
XX US6416959-B1.
XX
XX 09-JUL-2002.
XX
XX 25-FEB-2000; 2000US-00513783.
XX
XX 27-FEB-1997; 97US-00810983.
XX 27-FEB-1998; 98US-00031271.
XX 26-FEB-1999; 99US-0122152P.
XX 08-MAR-1999; 99US-0123399P.
XX 12-JUN-1999; 99US-00352171.
XX 31-AUG-1999; 99US-0151797P.
XX
PR
```

---

```
PR 17-SEP-1999; 99US-00398965.
PR 29-OCT-1999; 99US-00430656.
XX 01-DEC-1999; 99US-0168408P.
XX
PA (GIUL/) GIULIANO K.
XX (KAFU/) KAFUR R.
XX
XX Giuliano K, Kapur R;
XX
XX WPI; 2002-634730/68.
DR N-PSDB; ABS71491.
XX
XX Automated cell-based toxin detection, classification, and/or
XX identification by treating cells involves use of three classes of
XX luminescent reporter molecules such as detectors, classifiers or
XX identifiers.
XX
XX Example 10; Fig 29A; 214pp; English.
XX
XX The invention describes methods of automated detection, classification
XX and identification comprising treating cells containing luminescent
XX reporter molecules (I) in array of locations with a test substance, where
XX (I) are detectors, classifiers or identifiers, imaging cells in each
XX location to obtain luminescent signals and converting optical information
XX into digital data to interpret presence of toxins in the test substance.
XX The method are useful for detection of toxins chosen from proteases, ADP-
XX ribosylating toxins, cytotoxic phospholipases, and exfoliative toxins.
XX Three classes of cell-based luminescent reporter molecules such as
XX detectors, classifiers and identifiers are described and serve as
XX reporters of toxic threat agents. The first two levels of
XX characterisation enable a rapid readout of toxin class without
XX sacrificing the ability to detect many new mutant toxins or dissect
XX several complex mixtures of known toxins. This is the amino acid sequence
XX of a protease biosensor related signal sequence used in the cell-based
XX screening system
XX
XX Sequence 239 AA;
SQ
    Query Match      98.5%; Score 1256; DB 5; Length 239;
    Best Local Similarity 98.7%; Pred. No. 1.8e-121;
    Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
QY 61 LVTGLSYGVCFSRYPDHMKQHDFFKSAPEGVYQVETIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQVETIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
RESULT 14
AAE14599
ID AAE14599 standard; protein; 239 AA.
XX
XX AAE14599;
XX
XX 31-MAY-2002 (first entry)
XX
XX Aequorea victoria enhanced green fluorescent protein.
DE
XX Mutagenesis; enhanced green fluorescent protein; EGFP; mutant; mutein.
XX
XX Aequorea victoria.
OS
XX Synthetic.
OS
```



XX FH Key Location/Qualifiers  
 FT Misc-difference 1. .3 /note= "Wild-type GFP Met-Ser are replaced with Met-Val-Ser"  
 FT FT Misc-difference 65 /note= "GFP Phe64 is replaced by Leu"  
 FT FT Misc-difference 66 /note= "GFP Ser65 is replaced by Thr"  
 FT XX EPl178109-A1.  
 XX XX 06-FEB-2002.  
 XX XX 03-AUG-2001; 2001EP-00306650.  
 XX PR 04-AUG-2000; 2000JP-00237166.  
 XX XX (RIKE ) RIKEN KK.  
 XX PI Miyawaki A, Sawano A;  
 XX PI WPI: 2002-208112/27.  
 XX DR N-PSDB; AAD27910.  
 XX  
 XX Method for mutagenesis, e.g. for introducing certain or random mutations at certain sites of the nucleotide sequence, comprises synthesizing a mutated strand and a complementary strand by use of megaprimers.  
 XX PS Example 1; Page 13-14; 31pp; English.  
 XX  
 XX The invention relates to a method for mutagenesis that comprises synthesizing a mutated strand and a complementary strand by use of megaprimers. The method basically comprises a DNA synthesis in which one or more primers that have a nucleotide sequence containing at least one mutation and a phosphorylated 5'-terminus are annealed to a template DNA and then subjected to an elongation reaction using a thermostable high-fidelity DNA polymerase, after which the phosphorylated 5'-terminus and the elongated terminus are ligated by means of a thermostable DNA ligase to synthesise a circular DNA containing the primers; a digestion in which the step of DNA synthesis is repeated several times to amplify the DNA containing the primers and then, at least DNAs other than the amplified circular DNA are digested into several fragments; and a double-stranded DNA synthesis in which, with the several fragments obtained in the step of digestion as megaprimers, the megaprimers are annealed to the circular DNA synthesised above, followed by an elongation reaction performed using the thermostable high-fidelity DNA polymerase. The method is useful for mutagenesis, particularly for introducing certain mutations at certain sites of a nucleotide sequence, or for introducing random mutations at certain sites of the nucleotide sequence. The present method is simple, speedy, economical and widely applicable. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used for mutagenesis in an exemplification of the invention. The EGFP is derived by altering the green fluorescent protein (GFP) sequence of Aequorea victoria  
 XX SQ Sequence 239 AA;  
 Query Match 98.5%; Score 1256; DB 5; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 1.8e-121;  
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60  
 DB 1 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60  
 QY 61 LVTGLSYGVCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDYL 120  
 DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDYL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 QY 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
 DB 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

QY 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
 DB 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
 RESULT 15  
 AAE34958  
 ID AAE34958 standard; protein; 239 AA.  
 XX AC AAE34958;  
 XX DT 28-MAY-2003 (first entry)  
 XX DE Aequorea victoria enhanced green fluorescent protein (EGFP).  
 XX KW Phosphorylation indicator; fluorescent protein; detection; phosphatase;  
 XX KW kinase; enhanced green fluorescent protein; EGFP.  
 XX OS Aequorea victoria.  
 XX PN WO200295058-A2.  
 XX PD 28-NOV-2002.  
 XX PF 24-MAY-2002; 2002WO-US016955.  
 XX PR 24-MAY-2001; 2001US-00865291.  
 XX PA (REGC ) UNIV CALIFORNIA.  
 XX PI Tsien RY, Ting AY, Zhang J;  
 XX WPI: 2003-148474/14.  
 XX DR N-PSDB; AA553428.  
 XX PT Novel chimeric phosphorylation indicators, useful for detecting kinase/phosphatase in samples, has donor molecule, phosphorylatable domain, phosphoaminoacid binding domain, and acceptor molecule, in operative linkage.  
 XX PS Disclosure; Col 56-57; 38pp; English.  
 XX The present invention relates to chimeric phosphorylation indicators comprising a phosphorylation polypeptide and a fluorescent protein or in operative linkage, a donor molecule, a phosphorylatable domain, a phosphoaminoacid binding domain (PAABD) and an acceptor molecule. The phosphorylation indicators of the invention are useful for detecting kinases or phosphatases in a biological sample. They are also useful in high throughput analysis e.g. for detecting a kinase inhibitor or phosphatase inhibitor. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used in the invention  
 XX SQ Sequence 239 AA;  
 Query Match 98.5%; Score 1256; DB 6; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 1.8e-121;  
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60  
 DB 1 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60  
 QY 61 LVTGLSYGVCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDYL 120  
 DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDYL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 QY 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
 DB 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

Db 131 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKEDHNVLLFEFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 15:54:08  
Job time : 50.1111 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:55 ; Search time 12.7778 Seconds  
(without alignments)  
965.630 Million cell updates/sec

Title: US-09-887-784-64G  
Perfect score: 1275  
Sequence: 1 MWSKGBELFTGVVPILVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patente AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pdp.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pdp.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pdp.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pdp.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS COMB.pdp.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pdp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1256	98.5	239	US-09-172-063-3	Sequence 3, Appli
2	1256	98.5	239	US-09-513-783A-46	Sequence 46, Appli
3	1256	98.5	239	US-09-316-919-4	Sequence 4, Appli
4	1256	98.5	239	US-09-602-641-3	Sequence 3, Appli
5	1256	98.5	239	US-09-920-922-2	Sequence 2, Appli
6	1256	98.5	281	US-09-062-102-1	Sequence 1, Appli
7	1256	98.5	281	US-09-364-946-1	Sequence 1, Appli
8	1256	98.5	294	US-09-513-783A-2	Sequence 2, Appli
9	1256	98.5	323	US-09-172-063-21	Sequence 21, Appli
10	1256	98.5	323	US-09-602-641-21	Sequence 21, Appli
11	1256	98.5	364	US-09-085-305-6	Sequence 6, Appli
12	1256	98.5	379	US-09-417-197-129	Sequence 129, App
13	1256	98.5	434	US-09-800-170-48	Sequence 48, Appli
14	1256	98.5	442	US-09-417-197-127	Sequence 127, App
15	1256	98.5	459	US-09-513-783A-170	Sequence 170, App
16	1256	98.5	544	US-09-417-197-113	Sequence 113, App
17	1256	98.5	544	US-09-417-197-115	Sequence 115, App
18	1256	98.5	604	US-09-417-197-59	Sequence 59, Appli
19	1256	98.5	605	US-09-417-197-41	Sequence 41, Appli
20	1256	98.5	606	US-09-417-197-65	Sequence 65, Appli
21	1256	98.5	607	US-09-417-197-47	Sequence 47, Appli
22	1256	98.5	630	US-09-417-197-63	Sequence 63, Appli
23	1256	98.5	631	US-09-417-197-39	Sequence 39, Appli
24	1256	98.5	633	US-09-417-197-45	Sequence 45, Appli
25	1256	98.5	635	US-09-417-197-125	Sequence 125, App
26	1256	98.5	642	US-08-818-253-2	Sequence 2, Appli
27	1256	98.5	642	US-08-818-253-6	Sequence 6, Appli

Sequence 2, Appli  
Sequence 6, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 75, Appli  
Sequence 51, Appli  
Sequence 71, Appli  
Sequence 139, App  
Sequence 176, App  
Sequence 141, App  
Sequence 143, App  
Sequence 77, Appli  
Sequence 178, App  
Sequence 53, Appli  
Sequence 61, Appli  
Sequence 43, Appli  
Sequence 117, App  
Sequence 119, App

ALIGNMENTS

RESULT 1  
US-09-172-063-3  
; Sequence 3, Application US/09172063  
; Patent No. 6150176  
; GENERAL INFORMATION:  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Miyawaki, Atsushi  
; APPLICANT: Liopis, Juan  
; APPLICANT: Wachter, Rebekka M.  
; APPLICANT: Remington, S. James  
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
; FILE REFERENCE: 07257/071001  
; CURRENT APPLICATION NUMBER: US/09/172,063  
; CURRENT FILING DATE: 1998-10-13  
; EARLIER APPLICATION NUMBER: 09/094,359  
; EARLIER FILING DATE: 1998-06-09  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: EGFP  
US-09-172-063-3

Query Match 98.5%; Score 1256; DB 3; Length 239;  
Best Local Similarity 98.7%; Pred. No. 1.1e-127;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MWSKGBELFTGVVPILVELDGVNCHKFSVSGEGDATYKLTLPKICTTGKLPVWPWT 60  
DB 1 MWSKGBELFTGVVPILVELDGVNCHKFSVSGEGDATYKLTLPKICTTGKLPVWPWT 60  
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNVKTAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNVKTAEVKFEGDTL 120  
QY 121 VNRIELKGDIFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFIRNIEDGSVOLA 180  
DB 121 VNRIELKGDIFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFIRNIEDGSVOLA 180  
QY 181 DHYQQNTPIGDGPVLLPDNHNHLYSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQQNTPIGDGPVLLPDNHNHLYSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

```
RESULT 2
US-09-513-793A-46
; Sequence 46, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kaput, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGFP
US-09-513-783A-46

Query Match          98.5%; Score 1256; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.1e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTLKFICTTGKLPVPWPT 60

QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRLEKGDIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 3
US-09-316-919-4
; Sequence 4, Application US/09316919
; Patent No. 6469154
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: 07257/073001
; CURRENT APPLICATION NUMBER: US/09/316,919
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-316-919-4

Query Match          98.5%; Score 1256; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.1e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTLKFICTTGKLPVPWPT 60

QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRLEKGDIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 4
US-09-602-641-3
; Sequence 3, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (O)...(O)
; OTHER INFORMATION: EGFP
US-09-602-641-3

Query Match          98.5%; Score 1256; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.1e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTLKFICTTGKLPVPWPT 60

QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRLEKGDIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 5
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. 6673610
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; 
```

[illegible]

```
Db 1 MVSKEELFTGVVPIILVELDGVNKHFSVSGEGDATYKGLTLKFKICTTGKLPVWPPT 60
QY 51 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
Db 51 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
QY 121 VNRLEKIDGDFKEDGNILGHKLEYNVSHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 121 VNRLEKIDGDFKEDGNILGHKLEYNVSHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 9
US-09-172-063-21
; Sequence 21, Application US/09172063
; Patent No. 6150176
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: GT-EGFP
US-09-172-063-21

Query Match 98.5%; Score 1256; DB 3; Length 323;
Best Local Similarity 98.7%; Pred. No. 1.7e-127; Indels 0; Gaps 0;
Matches 236; Conservative 1; Mismatches 2;

QY 1 MVSKEELFTGVVPIILVELDGVNKHFSVSGEGDATYKGLTLKFKICTTGKLPVWPPT 60
Db 85 MVSKEELFTGVVPIILVELDGVNKHFSVSGEGDATYKGLTLKFKICTTGKLPVWPPT 144
QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
Db 145 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 204
QY 121 VNRLEKIDGDFKEDGNILGHKLEYNVSHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 205 VNRLEKIDGDFKEDGNILGHKLEYNVSHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 264
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 265 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 323

RESULT 10
US-09-602-641-21
; Sequence 21, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
```

```
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; PRIOR FILING DATE: 2000-06-22
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: GT-EGFP
US-09-602-641-21

Query Match 98.5%; Score 1256; DB 4; Length 323;
Best Local Similarity 98.7%; Pred. No. 1.7e-127; Indels 0; Gaps 0;
Matches 236; Conservative 1; Mismatches 2;

QY 1 MVSKEELFTGVVPIILVELDGVNKHFSVSGEGDATYKGLTLKFKICTTGKLPVWPPT 60
Db 85 MVSKEELFTGVVPIILVELDGVNKHFSVSGEGDATYKGLTLKFKICTTGKLPVWPPT 144
QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
Db 145 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 204
QY 121 VNRLEKIDGDFKEDGNILGHKLEYNVSHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 205 VNRLEKIDGDFKEDGNILGHKLEYNVSHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 264
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 265 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 323

RESULT 11
US-09-085-305-6
; Sequence 6, Application US/09085305
; Patent No. 6191269
; GENERAL INFORMATION:
; APPLICANT: Pollock, Allan
; APPLICANT: Lovett, David H.
; APPLICANT: Turck, Johanna
; TITLE OF INVENTION: Selective Induction of Apoptosis in
; TITLE OF INVENTION: Malignant Cancer Cells by Delivery of N-Terminal
; TITLE OF INVENTION: Interleukin-1-Alpha Pro-Piece Polypeptide
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; STREET: 285 Hamilton Ave, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,305
; FILING DATE: 29-MAY-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
```

ATTORNEY/AGENT INFORMATION:  
NAME: Francis, Carol L  
REGISTRATION NUMBER: 36,513  
REFERENCE/DOCKET NUMBER: 6510/102US1  
TELEPHONE: 650-327-3400  
TELEFAX: 650-327-3231  
TELEX:  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 364 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-085-305-6

Query Match 98.5%; Score 1256; DB 3; Length 364;  
Best Local Similarity 98.7%; Pred. No. 2.1e-127;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYVKLTLPFICTTGLKLPVPWPT 60  
DB 126 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYVKLTLPFICTTGLKLPVPWPT 185

QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 186 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 245

QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVOLA 180  
DB 246 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVOLA 305

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
DB 306 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 364

RESULT 12  
US-09-417-197-129  
Sequence 129, Application US/09417197  
Patent No. 6518021  
GENERAL INFORMATION:  
APPLICANT: Ole Thastrup, et al.  
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An  
FILE REFERENCE: 3759-0110P  
CURRENT FILING DATE: 1999-10-07  
NUMBER OF SEQ ID NOS: 143  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 129  
LENGTH: 379  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: actin-binding-domain-EGFP fusion  
US-09-417-197-129

Query Match 98.5%; Score 1256; DB 4; Length 379;  
Best Local Similarity 98.7%; Pred. No. 2.2e-127;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYVKLTLPFICTTGLKLPVPWPT 60  
DB 141 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYVKLTLPFICTTGLKLPVPWPT 200

QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 201 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 260

QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVOLA 180  
DB 261 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVOLA 320

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
DB 321 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 379

RESULT 13  
US-09-800-170-48  
Sequence 48, Application US/09800170  
Patent No. 6481667  
GENERAL INFORMATION:  
APPLICANT: Kinsella, Todd  
TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES  
FILE REFERENCE: A-68614-1/DJB/RMS/RMK  
CURRENT APPLICATION NUMBER: US/09/800,170  
CURRENT FILING DATE: 2001-08-28  
PRIOR APPLICATION NUMBER: US 60/187,130  
PRIOR FILING DATE: 2000-03-06  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 48  
LENGTH: 434  
TYPE: PRT  
ORGANISM: Synechocystis PCC6803  
US-09-800-170-48

Query Match 98.5%; Score 1256; DB 4; Length 434;  
Best Local Similarity 98.7%; Pred. No. 2.7e-127;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYVKLTLPFICTTGLKLPVPWPT 60  
DB 196 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYVKLTLPFICTTGLKLPVPWPT 255

QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 256 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 315

QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVOLA 180  
DB 316 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVOLA 375

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
DB 376 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 434

RESULT 14  
US-09-417-197-127  
Sequence 127, Application US/09417197  
Patent No. 6518021  
GENERAL INFORMATION:  
APPLICANT: Ole Thastrup, et al.  
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An  
FILE REFERENCE: 3759-0110P  
CURRENT FILING DATE: 1999-10-07  
NUMBER OF SEQ ID NOS: 143  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 127  
LENGTH: 442  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: EGFP-RhoA fusion  
US-09-417-197-127

Query Match 98.5%; Score 1256; DB 4; Length 442;  
Best Local Similarity 98.7%; Pred. No. 2.8e-127;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYVKLTLPFICTTGLKLPVPWPT 60

Db	1	MVSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFTCTTGKLPVPWPT	60
Qy	61	LVTGLSYGVCFSRYPDHMKQHDFFKSAWPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL	120
Db	61	LVTTLTYGVCFSRYPDHMKQHDFFKSAWPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL	120
Qy	121	VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA	180
Db	121	VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA	180
Qy	181	DHYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLLGFVTAAGITLGMDELYK	239
Db	181	DHYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLLGFVTAAGITLGMDELYK	239

RESULT 15  
US-09-513-783A-170  
; Sequence 170, Application US/09513783A  
; Patent No. 6416959  
; GENERAL INFORMATION:  
; APPLICANT: Giuliano, Kenneth A.  
; TITLE OF INVENTION: A System for Cell Based Screening  
; FILE REFERENCE: 97-022-L1  
; CURRENT APPLICATION NUMBER: US/09/513,783A  
; CURRENT FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 170  
; LENGTH: 459  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP27  
US-09-513-783A-170

Query Match	98.5%	Score 1256;	DB 4;	Length 459;
Best Local Similarity	98.7%	Pred. No. 2.9e-127;	Indels 0;	Gaps 0;
Matches 236;	Conservative 1;	Mismatches 2;		
Qy	1	MVSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFTCTTGKLPVPWPT	60	
Db	1	MVSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFTCTTGKLPVPWPT	60	
Qy	61	LVTGLSYGVCFSRYPDHMKQHDFFKSAWPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL	120	
Db	61	LVTTLTYGVCFSRYPDHMKQHDFFKSAWPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL	120	
Qy	121	VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA	180	
Db	121	VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA	180	
Qy	181	DHYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLLGFVTAAGITLGMDELYK	239	
Db	181	DHYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLLGFVTAAGITLGMDELYK	239	

Search completed: June 21, 2004, 16:04:00  
Job time : 13.7778 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:54:20 ; Search time 34.7778 Seconds  
(without alignments)  
1940.117 Million cell updates/sec

Title: US-09-887-784-64G  
Perfect score: 1275  
Sequence: 1 MVSKEELFTGVVPILVELD.....VLLGFVTRAGITLGMDELYK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1267	99.4	239	9	US-09-887-784-4
2	1267	99.4	239	12	US-10-296-953-4
3	1267	99.4	363	14	US-10-270-223-6
4	1267	99.4	893	14	US-10-257-909A-30
5	1267	99.4	1132	14	US-10-257-909A-32
6	1259	98.7	239	9	US-09-887-784-2
7	1259	98.7	239	12	US-10-296-953-2
8	1256	98.5	239	9	US-09-920-922-2
9	1256	98.5	239	9	US-09-999-745-4
10	1256	98.5	239	10	US-09-866-538-4
11	1256	98.5	239	10	US-09-797-496B-2
12	1256	98.5	239	10	US-09-794-308-4
13	1256	98.5	239	10	US-09-865-291-4
14	1256	98.5	239	12	US-10-457-982-3
15	1256	98.5	239	14	US-10-121-258-13

16	1256	98.5	239	14	US-10-221-461-7	Sequence 7, Appli
17	1256	98.5	239	14	US-10-100-957A-46	Sequence 46, Appli
18	1256	98.5	239	14	US-10-177-390-2	Sequence 2, Appli
19	1256	98.5	239	14	US-10-338-411-3	Sequence 3, Appli
20	1256	98.5	239	15	US-10-370-570-4	Sequence 4, Appli
21	1256	98.5	239	15	US-10-389-640-3	Sequence 3, Appli
22	1256	98.5	259	14	US-10-314-861-11	Sequence 11, Appli
23	1256	98.5	281	12	US-09-931-232-1	Sequence 1, Appli
24	1256	98.5	288	14	US-10-314-861-37	Sequence 37, Appli
25	1256	98.5	293	14	US-10-314-861-35	Sequence 35, Appli
26	1256	98.5	294	14	US-10-100-957A-2	Sequence 2, Appli
27	1256	98.5	295	14	US-10-314-861-39	Sequence 39, Appli
28	1256	98.5	299	14	US-10-314-861-33	Sequence 33, Appli
29	1256	98.5	305	14	US-10-314-861-31	Sequence 31, Appli
30	1256	98.5	308	14	US-10-033-717-35	Sequence 35, Appli
31	1256	98.5	311	14	US-10-314-861-29	Sequence 29, Appli
32	1256	98.5	320	14	US-10-338-411-11	Sequence 11, Appli
33	1256	98.5	320	15	US-10-389-640-11	Sequence 11, Appli
34	1256	98.5	323	12	US-10-457-982-21	Sequence 21, Appli
35	1256	98.5	323	14	US-10-338-411-7	Sequence 7, Appli
36	1256	98.5	323	14	US-10-338-411-13	Sequence 13, Appli
37	1256	98.5	323	15	US-10-389-640-7	Sequence 7, Appli
38	1256	98.5	323	15	US-10-389-640-13	Sequence 13, Appli
39	1256	98.5	324	14	US-10-314-861-16	Sequence 16, Appli
40	1256	98.5	345	14	US-10-338-411-5	Sequence 5, Appli
41	1256	98.5	345	15	US-10-389-640-5	Sequence 5, Appli
42	1256	98.5	346	14	US-10-338-411-9	Sequence 9, Appli
43	1256	98.5	346	15	US-10-389-640-9	Sequence 9, Appli
44	1256	98.5	359	14	US-10-033-717-33	Sequence 33, Appli
45	1256	98.5	359	14	US-10-033-717-34	Sequence 34, Appli

ALIGNMENTS

RESULT 1

US-09-887-784-4  
; Sequence 4, Application US/09887784  
; Patent No. US20020177189A1  
; GENERAL INFORMATION:  
; APPLICANT: BIORN, Sara et al  
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS  
; FILE REFERENCE: 3759-0115P  
; CURRENT APPLICATION NUMBER: US/09/887,784  
; CURRENT FILING DATE: 2001-06-19  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequoria Victoria  
US-09-887-784-4

Query Match 99.4%; Score 1267; DB 9; Length 239;  
Best Local Similarity 99.6%; Pred. No. 1,1e-123;  
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGKLT	KFICTTGKLPVWPPT 60
Db	1	MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGKLT	KFICTTGKLPVWPPT 60
Qy	61	LVTGLSYGVQCFSRYPDHMKQDFFKSA	MPGVQERTIFFKDDGNYKTRAEVKFEGDYL 120
Db	61	LVTGLSYGVQCFSRYPDHMKQDFFKSA	MPGVQERTIFFKDDGNYKTRAEVKFEGDYL 120
Qy	121	VNRIELKIDFDKEDGNILGHKLEYNHN	YIMADKQKNGIKVNFIRHNIEDGSVQLA 180
Db	121	VNRIELKIDFDKEDGNILGHKLEYNHN	YIMADKQKNGIKVNFIRHNIEDGSVQLA 180
Qy	181	DHYQONTPIGDGFPVLLPDNHYLSTQ	SALSKDNEKRDHNVLLGFVTAAGITLGMDELYK 239
Db	181	DHYQONTPIGDGFPVLLPDNHYLSTQ	SALSKDNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 2  
US-10-296-953-4  
; Sequence 4, Application US/10296953  
; Publication No. US20040072995A1  
; GENERAL INFORMATION:  
; APPLICANT: BJORN, SARA P.  
; APPLICANT: PAGLIARO, LEN  
; APPLICANT: THASTRUP, OLE  
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS  
; FILE REFERENCE: PLO095  
; CURRENT APPLICATION NUMBER: US/10/296,953  
; CURRENT FILING DATE: 2002-11-26  
; PRIOR APPLICATION NUMBER: PA 2000 00953  
; PRIOR FILING DATE: 2000-06-19  
; PRIOR APPLICATION NUMBER: 60/212,681  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: 60/290,170  
; PRIOR FILING DATE: 2001-05-10  
; PRIOR APPLICATION NUMBER: PA 2001 00739  
; PRIOR FILING DATE: 2001-05-10  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
US-10-296-953-4  
Query Match 99.4%; Score 1267; DB 12; Length 239;  
Best Local Similarity 99.6%; Pred. No. 1.1e-123; Indels 0; Gaps 0;  
Matches 238; Conservative 0; Mismatches 1;  
1 MVSKEELFTGVVPIVVELDGVNGHKFSVSGEGDATYKGLTLKFKICTTGKLPVWPPT 60  
1 MVSKEELFTGVVPIVVELDGVNGHKFSVSGEGDATYKGLTLKFKICTTGKLPVWPPT 60  
61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239  
181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239  
Query Match 99.4%; Score 1267; DB 12; Length 239;  
Best Local Similarity 99.6%; Pred. No. 1.1e-123; Indels 0; Gaps 0;  
Matches 238; Conservative 0; Mismatches 1;  
1 MVSKEELFTGVVPIVVELDGVNGHKFSVSGEGDATYKGLTLKFKICTTGKLPVWPPT 60  
1 MVSKEELFTGVVPIVVELDGVNGHKFSVSGEGDATYKGLTLKFKICTTGKLPVWPPT 60  
61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239  
181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239  
Query Match 99.4%; Score 1267; DB 14; Length 363;  
Best Local Similarity 99.6%; Pred. No. 1.9e-123; Indels 0; Gaps 0;  
Matches 238; Conservative 0; Mismatches 1;  
US-10-270-423-6  
; Sequence 6, Application US/10270223  
; Publication No. US20030143634A1  
; GENERAL INFORMATION:  
; APPLICANT: BioImage A/S  
; TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR COMPONENTS  
; TITLE OF INVENTION: INTACT LIVING CELLS, AND TO EXTRACT QUANTITATIVE INFORMATION RE  
; TITLE OF INVENTION: INTERACTIONS BY FLUORESCENCE REDISTRIBUTION.  
; FILE REFERENCE: 3759-0126P  
; CURRENT APPLICATION NUMBER: US/10/270,223  
; CURRENT FILING DATE: 2002-10-11  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 6  
; LENGTH: 363  
; TYPE: PRT  
; ORGANISM: Aequorea Victoria and Human  
US-10-270-423-6  
Query Match 99.4%; Score 1267; DB 14; Length 363;  
Best Local Similarity 99.6%; Pred. No. 1.9e-123; Indels 0; Gaps 0;  
Matches 238; Conservative 0; Mismatches 1;  
US-10-257-909A-32  
; Sequence 32, Application US/10257909A  
; Publication No. US20030187056A1  
; GENERAL INFORMATION:  
; APPLICANT: Bernard R. TERRY et al.  
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular  
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes  
; FILE REFERENCE: 3759-0125P  
; CURRENT APPLICATION NUMBER: US/10/257,909A  
; CURRENT FILING DATE: 2002-10-17  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 30  
; LENGTH: 893  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Fusion between Aequorea victoria and human  
US-10-257-909A-30  
Query Match 99.4%; Score 1267; DB 14; Length 893;  
Best Local Similarity 99.6%; Pred. No. 7e-123; Indels 0; Gaps 0;  
Matches 238; Conservative 0; Mismatches 1;  
1 MVSKEELFTGVVPIVVELDGVNGHKFSVSGEGDATYKGLTLKFKICTTGKLPVWPPT 60  
655 MVSKEELFTGVVPIVVELDGVNGHKFSVSGEGDATYKGLTLKFKICTTGKLPVWPPT 714  
61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
715 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 774  
121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
775 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 834  
181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239  
835 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 893  
RESULT 5  
US-10-257-909A-32  
; Sequence 32, Application US/10257909A  
; Publication No. US20030187056A1  
; GENERAL INFORMATION:  
; APPLICANT: Bernard R. TERRY et al.  
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular  
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes  
; FILE REFERENCE: 3759-0125P  
; CURRENT APPLICATION NUMBER: US/10/257,909A  
; CURRENT FILING DATE: 2002-10-17  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 3.0

QY 1 MVSKEELFTGVVPIVVELDGVNGHKFSVSGEGDATYKGLTLKFKICTTGKLPVWPPT 60  
DB 1 MVSKEELFTGVVPIVVELDGVNGHKFSVSGEGDATYKGLTLKFKICTTGKLPVWPPT 60  
QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239  
RESULT 4  
US-10-257-909A-30  
; Sequence 30, Application US/10257909A  
; Publication No. US20030187056A1  
; GENERAL INFORMATION:  
; APPLICANT: Bernard R. TERRY et al.  
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular  
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes  
; FILE REFERENCE: 3759-0125P  
; CURRENT APPLICATION NUMBER: US/10/257,909A  
; CURRENT FILING DATE: 2002-10-17  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 30  
; LENGTH: 893  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Fusion between Aequorea victoria and human  
US-10-257-909A-30  
Query Match 99.4%; Score 1267; DB 14; Length 893;  
Best Local Similarity 99.6%; Pred. No. 7e-123; Indels 0; Gaps 0;  
Matches 238; Conservative 0; Mismatches 1;  
1 MVSKEELFTGVVPIVVELDGVNGHKFSVSGEGDATYKGLTLKFKICTTGKLPVWPPT 60  
655 MVSKEELFTGVVPIVVELDGVNGHKFSVSGEGDATYKGLTLKFKICTTGKLPVWPPT 714  
61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
715 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 774  
121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
775 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 834  
181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239  
835 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 893

```
; SEQ ID NO 32
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-32

Query Match          99.4%; Score 1267; DB 14; Length 1132;
Best Local Similarity 99.6%; Pred. No. 9.8e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVWPWT 60
Db 894 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVWPWT 953

QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 954 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 1013

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 1014 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 1073

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLGVFTAAGITLGMDELYK 239
Db 1074 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLGVFTAAGITLGMDELYK 1132

RESULT 6
US-09-887-784-2
; Sequence 2, Application US/09887784
; Patent No. US20020177189A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea Victoria
US-09-887-784-2

Query Match          98.7%; Score 1259; DB 9; Length 239;
Best Local Similarity 99.2%; Pred. No. 7.2e-123;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVWPWT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVWPWT 60

QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLGVFTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLGVFTAAGITLGMDELYK 239

RESULT 7
US-10-296-953-2
; Sequence 2, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PLO095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match          98.5%; Score 1256; DB 9; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVWPWT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVWPWT 60
```

```
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PLO095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match          98.7%; Score 1259; DB 12; Length 239;
Best Local Similarity 99.2%; Pred. No. 7.2e-123;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVWPWT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVWPWT 60

QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLGVFTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLGVFTAAGITLGMDELYK 239

RESULT 8
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. US20020083488A1
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match          98.5%; Score 1256; DB 9; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVWPWT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVWPWT 60
```

```
QY 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 9
US-09-999-745-4
; Sequence 4, Application US/09999745
; Patent No. US20020157120A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGEN1470-1
; CURRENT APPLICATION NUMBER: US/09/999,745
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
; PRIOR FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-999-745-4

Query Match 98.5%; Score 1256; DB 9; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKLTLPICCTTGKLPVPWPT 60
1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKLTLPICCTTGKLPVPWPT 60
61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 10
US-09-866-538-4
; Sequence 4, Application US/09866538
; Patent No. US20030032088A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Campbell, Robert
; APPLICANT: Tsien, Roger
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/09/866,538
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 4
; LENGTH: 239
; TYPE: PRT
```

```
; ORGANISM: Aequorea victoria
US-09-866-538-4

Query Match 98.5%; Score 1256; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKLTLPICCTTGKLPVPWPT 60
1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKLTLPICCTTGKLPVPWPT 60
61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 11
US-09-797-496B-2
; Sequence 2, Application US/09797496B
; Publication No. US20030049597A1
; GENERAL INFORMATION:
; APPLICANT: Simon, Sanford M.
; APPLICANT: Chen, Yu
; TITLE OF INVENTION: Chimeric Fluorescent Enzymes and Uses Thereof
; FILE REFERENCE: 600-1-267
; CURRENT APPLICATION NUMBER: US/09/797,496B
; CURRENT FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Aequorea victoria green fluorescent protein modified as described
; OTHER INFORMATION: in specification
US-09-797-496B-2

Query Match 98.5%; Score 1256; DB 10; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKLTLPICCTTGKLPVPWPT 60
1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKLTLPICCTTGKLPVPWPT 60
61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 12
US-09-794-308-4
; Sequence 4, Application US/09794308
; Publication No. US20030170911A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
```

APPLICANT: TSIEH, Roger  
APPLICANT: ZACHARIAS, David  
APPLICANT: BAIRD, Geoffrey  
TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS  
FILE REFERENCE: REG1530  
CURRENT APPLICATION NUMBER: US/09/794,308  
CURRENT FILING DATE: 2001-02-26  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
US-09-794-308-4

Query Match 98.5%; Score 1256; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 1.5e-122;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHFKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHFKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
QY 61 LVTGLSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 13  
US-09-865-291-4  
Sequence 4, Application US/09865291  
Publication No. US20030186229A1  
GENERAL INFORMATION:  
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA  
APPLICANT: TSIEH, Roger  
APPLICANT: TING, Alice  
APPLICANT: ZHANG, Jin  
TITLE OF INVENTION: EMISSION RATIONETRIC INDICATORS OF PHOSPHORYLATION  
FILE REFERENCE: REG1550  
CURRENT APPLICATION NUMBER: US/09/865,291  
CURRENT FILING DATE: 2001-05-24  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
US-09-865-291-4

Query Match 98.5%; Score 1256; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 1.5e-122;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHFKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHFKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
QY 61 LVTGLSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 14  
US-10-457-982-3  
Sequence 3, Application US/10457982  
Publication No. US2003021265A1  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger Y.  
APPLICANT: Miyawaki, Acsushi  
APPLICANT: Llopis, Juan  
APPLICANT: Wichter, Rebekka M.  
APPLICANT: Remington, S. James  
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
FILE REFERENCE: 07257/071001  
CURRENT APPLICATION NUMBER: US/10/457,982  
PRIOR FILING DATE: 2003-06-09  
PRIOR APPLICATION NUMBER: US/09/602,641  
PRIOR FILING DATE: 2000-06-22  
PRIOR APPLICATION NUMBER: 09/172,063  
PRIOR FILING DATE: 1998-10-13  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (0)...(0)  
OTHER INFORMATION: EGFP  
US-10-457-982-3

Query Match 98.5%; Score 1256; DB 12; Length 239;  
Best Local Similarity 98.7%; Pred. No. 1.5e-122;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHFKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHFKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
QY 61 LVTGLSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 15  
US-10-121-258-13  
Sequence 13, Application US/10121258  
Publication No. US20030059835A1  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger  
APPLICANT: Campbell, Robert  
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT  
TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME  
FILE REFERENCE: UC083.1CP2CP1  
CURRENT APPLICATION NUMBER: US/10/121,258  
CURRENT FILING DATE: 2002-04-10  
PRIOR APPLICATION NUMBER: 09/794,308  
PRIOR FILING DATE: 2001-02-26  
PRIOR APPLICATION NUMBER: 09/866,538  
PRIOR FILING DATE: 2001-05-24

Sun Jun 27 18:27:36 2004

; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Enhanced Green Fluorescent Protein (EGFP)  
US-10-121-258-13

Query Match 98.5%; Score 1256; DB 14; Length 239;  
Best Local Similarity 98.7%; Pred. No. 1.5e-122;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYGLTKPICKTKGLPVPWPT 60  
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYGLTKPICKTKGLPVPWPT 60  
Qy 61 LVTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Qy 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180  
Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180  
Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:09:25  
Job time : 35.7778 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:00 ; Search time 10.3333 Seconds  
(without alignments)  
2224.817 Million cell updates/sec

Title: US-09-887-784-64G  
Perfect score: 1275  
Sequence: 1 MYSKGELFTGVVILVELD.....VLLGFTVTAAGITLGMDELYK 239  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1234	96.8	238	1 JQ1514	green-fluorescent
2	106	8.3	785	2 H72228	hypothetical prote
3	92.5	7.3	797	2 JC4078	protective surface
4	92.5	7.3	808	2 F64102	protective surface
5	92.5	7.3	941	2 S29043	cellulase (EC 3.2.
6	90.5	7.1	740	2 G95153	neuraminidase, pro
7	88.5	6.9	655	2 D83197	DNA topoisomerase
8	88.5	6.9	861	2 H64102	leucine-tRNA ligas
9	87.5	6.9	370	2 E70390	iron-sulfur cofact
10	87.5	6.9	2518	2 A12140	polyketide synthas
11	87.5	6.9	2573	2 D71614	hypothetical prote
12	87	6.8	752	1 KXRTC1	proprotein convert
13	86.5	6.8	887	2 E82590	leucyl-tRNA synthet
14	86	6.7	357	2 G81355	trna (uracil-5-)-m
15	86	6.7	578	1 I40794	dihydrolipoamide d
16	85.5	6.7	217	2 S35324	exp1 protein - Erw
17	85.5	6.7	889	2 JC5576	inter-alpha-trypsi
18	85	6.7	281	2 AD2052	hypothetical prote
19	85	6.7	439	2 JH0414	synaptogamin o-p65
20	84.5	6.6	613	2 A99552	oligoneuroptide
21	84.5	6.6	788	1 JDLVLH	DNA-directed DNA p
22	84.5	6.6	860	2 AC0582	leucyl-tRNA synthet
23	83.5	6.5	425	2 C97354	hypothetical prote
24	83.5	6.5	1134	2 A60234	IgA Fc receptor pr
25	83.5	6.5	1164	1 FCSOAG	IgA Fc receptor pr
26	83.5	6.5	1603	2 F89497	protein vit-5 [imp
27	83	6.5	632	2 T06586	DNA-binding protei
28	83	6.5	782	2 S55094	probable membrane
29	82.5	6.5	836	1 JDLVD	DNA-directed DNA p

RESULT 1  
JQ1514

green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)  
C;Species: Aequorea victoria  
C;Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 23-Mar-2001  
C;Accession: JQ1514; PQ0335; S48693; S51330; S51331  
R;Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.  
Gene 111, 229-233, 1992  
A;Title: Primary structure of the Aequorea victoria green-fluorescent protein.  
A;Reference number: JQ1514; MUID:92175527; PMID:1347277  
A;Accession: JQ1514  
A;Molecule type: DNA  
A;Residues: 1-107, S', 109-238 <PRA1>  
A;Cross-references: GB:M62654; NID:G155662; PIDN:AAA27722.1; PID:G155663  
A;Accession: JQ1514  
A;Molecule type: mRNA  
A;Residues: 1-99, 'P', 101-140, 'L', 142-218, 'V', 220-238 <PRA2>  
A;Cross-references: GB:M62653; NID:G155660; PIDN:AAA27721.1; PID:G155661  
A;Accession: PQ0335

ALIGNMENTS

30	82.5	6.5	1224	1 ERHUAH	coatmer complex a
31	82	6.4	353	2 E84941	imidazoleglycerol-
32	82	6.4	461	2 T06936	photosystem II chl
33	82	6.4	471	2 T27856	hypothetical prote
34	82	6.4	774	2 T39539	alpha-amylase homo
35	82	6.4	836	2 T42323	hypothetical prote
36	82	6.4	865	2 T11852	lipoxigenase (EC 1
37	81.5	6.4	263	2 S53488	water-stress-induc
38	81.5	6.4	310	2 S68225	synergohymenotrophi
39	81.5	6.4	400	2 H69222	molybdenum formylm
40	81.5	6.4	1603	1 VJRW5	vitellogenin vit-5
41	81.5	6.4	1983	2 G86643	hypothetical prote
42	81	6.4	336	2 C64468	hypothetical prote
43	81	6.4	433	2 A75125	hypothetical prote
44	81	6.4	449	2 E75394	hypothetical prote
45	81	6.4	515	2 C72344	propionyl-CoA carb

A;Reference number: A58953; MUID:98294543; PMID:9631087

A;Contents: annotation: X-ray crystallography, 1.9 angstroms  
C;Comment: This protein is excited by the photoprotein aequorin (see PIR:AQFNV) emitting  
C;Comment: The chromophore of this protein is formed by modification of Ser-dehydro-Tyr-  
A;Gene: GFP  
A;Introns: 69/3; 167/3  
A;Superfamily: green-fluorescent protein  
C;Keywords: chromoprotein; luminescence  
F;65-67/Cross-link: 5-imidazolinone (Ser-Gly) #status experimental  
F;66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match 96.8%; Score 1234; DB 1; Length 238;  
Best Local Similarity 96.6%; Pred. No. 6.4e-96;  
Matches 230; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSGEELFTGVVILVELDGVNGHKFSVSGEGDATYKGLTLKFICTGKLPVWPVL 61  
DB 1 MSKGEELFTGVVILVELDGVNGHKFSVSGEGDATYKGLTLKFICTGKLPVWPVL 60  
QY 62 VTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFPKDGNKYKTRAEVKFEGDTLV 121  
DB 61 VTFISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120  
QY 122 NRLEKIDPKEDGNILGHKLEYNNSHNYYIMADKQNGIKYKIRHNIEDGSVOLAD 181  
DB 121 NRLEKIDPKEDGNILGHKLEYNNSHNYYIMADKQNGIKYKIRHNIEDGSVOLAD 180  
QY 182 HYQONTPIGDPVLLPNHLYSTQSALSADPNKRDHMLLVFVTAAGITLGMDELYK 239  
DB 181 HYQONTPIGDPVLLPNHLYSTQSALSADPNKRDHMLLVFVTAAGITLGMDELYK 238

RESULT 2  
H72228  
hypothetical protein TM1624 - Thermotoga maritima (strain MSB8)  
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A;References: A72200; MUID:99287316; PMID:10360571  
A;Accession: H72228  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-785 <ARN>  
A;Cross-references: GB:AE001806; GB:AE000512; NID:g4982196; PIDN:AAD36691.1; PID:g498219  
A;Experimental source: strain MSB8  
C;Genetics:  
A;Gene: TM1624

Query Match 8.3%; Score 106; DB 2; Length 785;  
Best Local Similarity 20.2%; Pred. No. 0.75;  
Matches 47; Conservative 33; Mismatches 69; Indels 84; Gaps 8;

QY 3 SKGEELFTGVVILVELDGVNGHKFSVSGEGDATYKGLTLKFICTGKLPVWPVL 62  
DB 15 NEGFSFEGTVPGVQAD-----LVRKGLLPHPY----- 43  
QY 63 TGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFPKDGNKYKTRAEVKFEGDTLV 122  
DB 44 VGMN-----EDLFKEIDBEWIEYEFPEFKEDVKEGERVDLVFEGVDTL 88  
QY 123 RLEKIDPKEDGNILGHKLEYNNSHNYYIMADKQNGIKYKIRHNIEDGSVOLADH 182  
DB 89 DYVLTGVYL--GSTEDMFEYRFVDTLV-----KEKNLKVYIK-----SPIRVPKT 134  
QY 183 YQONTPIGDPVLLPNHLYSTQSALSADPNKRDHMLLVFVTAAGITLGM 235  
DB 135 LEQNYGVLGSP-----EDF-----IRGYIRKAQSYGWD 163

RESULT 3  
JC4078  
protective surface antigen D-15 precursor - Haemophilus influenzae (type b)  
C;Species: Haemophilus influenzae  
A;Variety: type b  
C;Date: 30-Jun-1995 #sequence\_revision 14-Jul-1995 #text\_change 29-Sep-1999  
C;Accession: JC4078  
R;Flack, F.S.; Loomore, S.; Chong, P.; Thomas, W.R.  
Gene 156, 97-99, 1995  
A;Title: The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus infl  
A;Reference number: JC4078; MUID:95255676; PMID:7737523  
A;Accession: JC4078  
A;Molecule type: DNA  
A;Residues: 1-797 <FLA>  
A;Cross-references: GB:U13961; NID:9537447; PIDN:AAA85645.1; PID:g537448  
A;Experimental source: type b  
C;Superfamily: protective surface antigen D-15  
C;Keywords: surface antigen  
F;1-19/Domain: signal sequence #status predicted <SIG>  
F;20-797/Product: protective surface antigen D-15 #status predicted <MAT>

Query Match 7.3%; Score 92.5; DB 2; Length 797;  
Best Local Similarity 22.3%; Pred. No. 10;  
Matches 49; Conservative 30; Mismatches 78; Indels 63; Gaps 11;

QY 64 GLSYGVQCFSRYPDHMKQHD-----FKSAMPEGYVQ-----RT 98  
DB 426 GIGYGTSGISYQASVKQDNFLGTGAANSIAGTKNDYGTSVNLGYTEPYFTKDGVS LGN 485  
QY 99 IFFKDDGNKYKTRAEVKFEGDTLVNRLEKIDPKEDGNI---LGH-----KLEYNYN 147  
DB 486 VFFENYDNSKSDTSSNYKRTYGSNVTL--GFPVNENNSYYVGLGHTYKISNFALEYN-- 542  
QY 148 SHNVIYIMADKQK-NGIKYKIRHNIEDGSVOLADHYQ-----NTPIGDGPVL 195  
DB 543 -RNLYIQSMKFKNGIKTN-----DPDFSGWNYSNLRGYPTKGVKASLG-GRVT 592  
QY 196 LP--DNHLYSTQSALSADPNKRDHMLLVFVTAAGITLG 233  
DB 593 IFGSDNKKYKLSADVQGFYPLDRDLHWVVSASAKASAGYANG 632

RESULT 4  
F64102  
protective surface antigen D-15 - Haemophilus influenzae (strain Rd KW20)  
C;Species: Haemophilus influenzae  
C;Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 18-Sep-1998  
C;Accession: F64102  
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J  
; D.M.; Brandon, R.C.; Pine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.  
Science 269, 496-512, 1995  
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,  
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A;Reference number: A64000; MUID:95350630; PMID:7542800  
A;Accession: F64102  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-808 <TIGR>  
A;Cross-references: GB:L42023; TIGR:HI0917  
C;Superfamily: protective surface antigen D-15  
C;Keywords: surface antigen

Query Match 7.3%; Score 92.5; DB 2; Length 808;  
Best Local Similarity 22.3%; Pred. No. 10;  
Matches 49; Conservative 30; Mismatches 78; Indels 63; Gaps 11;

QY 64 GLSYGVQCFSRYPDHMKQHD-----FKSAMPEGYVQ-----RT 98  
DB 439 GIGYGTSGISYQASVKQDNFLGTGAANSIAGTKNDYGTSVNLGYTEPYFTKDGVS LGN 498  
QY 99 IFFKDDGNKYKTRAEVKFEGDTLVNRLEKIDPKEDGNI---LGH-----KLEYNYN 147



```

Db 499 VFFENYDNSKSDTSNXYKRTTYGNSVTL-CFFPVNENNSYVGLGHTYNNKISNPALEYN-- 555
QY 148 SHNVYIMADKQK-NGIKVAFKIRHNIEDSGVQLADHVQO-----NTPIGDGPVL 195
Db 556 -RMLYIQSMKFKNGIKTN-----DFFSFGWYNSLNRGYFFTKGVKASLG-GRVT 605
QY 196 LP--DNHYLSTQSALSKDPNEKRDMVLLGFTVTAAGITLG 233
Db 606 IPGSDNKYKLSADVQGFVPLDRDHLWVVSASAKASAGYANG 645

RESULT 5
S29043
cellulase (EC 3.2.1.4) - Bacillus sp.
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus sp.
C:Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 01-Dec-2000
C:Accession: S29043; PC4404
R:Ozaki, K.; Shikata, S.; Kawai, S.; Ito, S.; Okamoto, K.
J. Gen. Microbiol. 136, 1327-1334, 1990
A:Title: Molecular cloning and nucleotide sequence of a gene for alkaline cellulase from
A:Reference number: S29043; MUID:91037937; PMID:2230718
A:Accession: S29043
A:Molecule type: DNA
A:Residues: 1-941 <OZA>
A:Cross-references: EMBL:M27420; NID:g142664; PIDN:AAA22304.1; PID:g142665
R:Shirai, T.; Yamane, T.; Hidaka, T.; Kuyama, K.; Suzuki, A.; Ashida, T.; Ozaki, K.; Ito
J. Biochem. 122, 683-685, 1997
A:Title: Crystallization and preliminary X-ray analysis of a truncated family A alkaline
A:Reference number: PC4404; MUID:98060488; PMID:9399567
A:Accession: PC4404
A:Molecule type: protein
A:Residues: 228-584 <SHI>
A:Experimental source: strain KSM-635
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A:Pathway: cellulose degradation
C:Superfamily: Bacillus sp. KSM-635 alkaline cellulase; S-layer repeat homology; Thermob
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:41-95/Domain: S-layer repeat homology <SLR1>
F:101-153/Domain: S-layer repeat homology <SLR2>
F:164-219/Domain: S-layer repeat homology <SLR3>
F:766-908/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA>

Query Match 7.3%; Score 92.5; DB 2; Length 941;
Best Local Similarity 20.3%; Pred. No. 13;
Matches 49; Conservative 34; Mismatches 60; Indels 99; Gaps 11;
QY 16 LVELDGVNGHKFVSVEGEGDATYGGKLTLLKFICTTGKLPVPWPTLVTLGLS-YGVQCFGR 74
Db 240 LVELNG-----QUTLAGE---DGT-----PVQLRGWSTHGLQWFG- 271
QY 75 YPDHMKQHDFFKSAPEGVQVQERTIFPKDQNGYKTRAEVKFEGDTLVNRIELKIDGDFKED 134
Db 272 --EIVNENAFVLSNDGWSNMIRLAMYIGENVATNPVK--DLVYEGIELA----- 319
QY 135 GNILGHLEYNVSHNVYIMADKQNGIKVNFIRHNIEDGSV-----QLADHYQQTPTIG 190
Db 320 -----FEHDMYIVDMH-----VHAPGDPRADYVSGYDFPEETADHYKDH----- 360
QY 191 DGPVLLPDNHYLSTQSALSQDPN-----EKRDHMYL 221
Db 361 -----PKNHYIWEELANPEPNNGPGPLTNDEKGEAVKEAYEPIVEMLEKGDNMIL 414
QY 222 LG 223
Db 415 VG 416

RESULT 6
G95153
neuraminidase, probable [imported] - Streptococcus pneumoniae (strain TIGR4)

```

```

C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: G95153
R:Tetrelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, I.
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: G95153
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-740 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75424.1; PID:g14972808; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP1326

Query Match 7.1%; Score 90.5; DB 2; Length 740;
Best Local Similarity 23.1%; Pred. No. 14;
Matches 59; Conservative 23; Mismatches 80; Indels 93; Gaps 14;
QY 13 VPILVELDGVNGHKFVSVEGEGDATYG-----KTLK 46
Db 291 IPVLTYFS---NGRVFS-----SIDARYGGTHDFLNKINIATSYSDNGKWTWKPLTLA 342
QY 47 FICTTGKLPVPWPTLVTLGLSYGVQCFSRYPDHM-----KQDFFKSAPEGVQVQERTIF 100
Db 343 F-DDFAPVPLEPREVGGDRDLQSGATYDSVIVEKQKQVLMFADVPAG-VSPREAT 400
QY 101 FKD-----DGNKYTRAEVKFEGDTLVN-----RIELKIDGDFKED 134
Db 401 RKDSGVKQIDGNYLK--LRKQGDYDNYTIRENGTVYDTRNRPTEFSVD-KNFGIKQN 457
QY 135 GNIL---GHKLYNNSHNVYIMADKQNGIKV-----FKIRHNIEDGSVQLADH 182
Db 458 GNYLTVEQYSVFSFENNKKTEY-----RNGTKVHMNIFYKDALFKVVPNTNYIAYISSNDH 511
QY 183 YQONTPIGDGPVLLP 197
Db 512 GSS----WSAPTLLP 522

RESULT 7
D83917
DNA topoisomerase IV subunit B BH2140 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: D83917
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hiran
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: D83917
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-655 <STO>
A:Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BAB05859.1; GSPDB:GN001
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2140
C:Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B

Query Match 6.9%; Score 88.5; DB 2; Length 655;
Best Local Similarity 21.9%; Pred. No. 17;
Matches 53; Conservative 37; Mismatches 95; Indels 57; Gaps 9;
QY 22 DVNGHK---FSVSGEGEGDAT---YGLTLTKFI-----CTTGKLPVPWP 59
Db 63 NVTIHKDQSVSRDEGRGMPTGMHKLKGPTEPVLTLVHAGKFGGGVATSGGLHGVGA 122
QY 60 TLVTGLSYGVQCFSRYPDHMKQHDFFKSAPEGVQER-----TIFPKDGG----- 105

```

Sun Jun 27 18:27:37 2004

Db 123 SVVNLSEWLVEIKRDGWVEQRPENGKGPSTTLEKKGKTRQTGTTHFKDPDPTVFSTT 182  
 QY 106 --NYKTRAEVKFGDGLVNRIELKGDIFKEDGNILGHKLEYNVSHNVYIMADK----- 157  
 Db 183 NFNVETLSERLEAAFLKGLKIELVDLRDDTKVEPH-YEDGKAPVEYINEDKETLHPV 241  
 QY 158 -----QKNGIKVNFKIRHNIEDGSVOLADHYQOQNTPIGDGPVLLPDNHYLSTQSALSADP 212  
 Db 242 VFPNGSGNIEIEFAQFN--DGYTENVLSPVNVVTKDG-----GTHELGAKTAMTRAV 294  
 QY 213 NE 214  
 Db 295 NE 296  
 RESULT 8  
 H64102  
 leucine-tRNA ligase (EC 6.1.1.4) - Haemophilus influenzae (strain Rd KW20)  
 N:Alternate names: leucyl-tRNA synthetase  
 C:Species: Haemophilus influenzae  
 C>Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 03-Jun-2002  
 C:Accession: H64102  
 R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.  
 ; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.  
 ; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
 Science 265, 496-512, 1995  
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Frazer, C.M.; Smith, H.O.; Venter,  
 A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
 A:Reference number: A64000; MUID:95350630; PMID:7542800  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-861 <TIGR>  
 A:Cross-references: GB:U32774; GB:L42023; NID:gi573942; PIDN:AAC22581.1; PID:gi573943; T  
 C:Genetics:  
 A:Gene: leuS  
 A:Superfamily: leucine-tRNA ligase  
 C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis  
 Query Match 6.9%; Score 88.5; DB 2; Length 861;  
 Best Local Similarity 24.1%; Pred. No. 25;  
 Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;  
 QY 50 TTGKLPVWPVTLTGLSYGVCFSRYPDHMKQHDFFKSAMPEGYVOERTIFFKD----- 103  
 Db 314 TGDKLLPI-WVANFVLMHYGTGAVMAVPAH-DQRDF-----EPAQKYSUPIKQVIAPLA 364  
 QY 104 DGNKTRAEVKFGDGLVNRIELKGDIFKEDGNILGHKLEYNVSHNVYIMADK-QKNGI 162  
 Db 365 DEIDLTKQAFVEHGKLVNSDEFQKNP--DQAFNG-----IADKLEKLV 408  
 QY 163 ---KVNFKIRH-----NIEDGSVOLADHYQOQNTPIGDGPVLLPDNHYL- 202  
 Db 409 GKRQVNYLRDQVSRQRYWYGAPIMLTLENGDVVPA-----PMEDLPILPEDVYMD 461  
 QY 203 STQSALSADPN 213  
 Db 462 GVKSPINADPN 472  
 RESULT 9  
 E70390  
 iron-sulfur cofactor synthesis protein nifs - Aquifex aeolicus  
 N:Contains: L-cysteine sulfurtransferase (EC 2.8.1.1-)  
 C:Species: Aquifex aeolicus  
 C>Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 07-Dec-1999  
 C:Accession: E70390  
 R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O  
 V.  
 Nature 392, 353-358, 1998  
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
 A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: E70390  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-370 <AQF>  
 A:Cross-references: GB:AB000720; NID:g2983529; PIDN:AAC07111.1; PID:g2983536; GB:AE000065;  
 A:Experimental source: strain VF5  
 C:Genetics:  
 A:Gene: nifs1  
 C:Superfamily: nitrogen fixation protein nifs  
 C:Keywords: phosphoprotein; pyridoxal phosphate; sulfurtransferase  
 F:195/Binding site: pyridoxal phosphate (lys) (covalent) #status predicted  
 F:318/Active site: Cys (cysteine persulfide intermediate) #status predicted  
 Query Match 6.9%; Score 87.5; DB 2; Length 370;  
 Best Local Similarity 25.4%; Pred. No. 9.9;  
 Matches 49; Conservative 30; Mismatches 91; Indels 23; Gaps 8;  
 QY 4 KGEELFTGW----PILVELD---GDVNGHKP-SVSGEG-----EGDATYKGLTLKICT 50  
 Db 164 KGVPLTTDAVQAIGKIPTELKNISVATPSGHKFAIKSGFLYISDEANYEPLIVGGQE 223  
 QY 51 TQKLP-----VPWPTLVTLGSLYGVOCFSRYPDHMKQ-HDFFKSAMPEGYVOERTIFFKD 104  
 Db 224 NGRSGTENVVGILSLAKALEIIVSNFSRYQEQKLRDLFENLLEA-LPDAQIVGKA 282  
 QY 105 GNYKTRAEV---KPEGDTLVNRIELKGDIFKEDGNILGHKLEYNVSHNVYIMADKQKNG 161  
 Db 283 ERSPSISVIMPKFPGAEIVNKLSEKGIYCTSGSACLSGEYEPNKHMLKMGFSQEKALRM 342  
 QY 162 IKVNFKIRHNIED 174  
 Db 343 VRFSFGLLNKEE 355  
 RESULT 10  
 AI2140  
 polyketide synthase [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp. PCC 7120  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
 C:Accession: AI2140  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,  
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takaizawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AI2140  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2518 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BA074379.1; PID:gi17131773; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: air2680  
 Query Match 6.9%; Score 87.5; DB 2; Length 2518;  
 Best Local Similarity 24.5%; Pred. No. 1.2e+02;  
 Matches 48; Conservative 24; Mismatches 69; Indels 55; Gaps 11;  
 QY 48 ICTTGKLPVWPVTLTGLSYGVCFSRY---PDH-----MKQHDFFK----- 86  
 Db 1537 IVLEGTIPSIVIDLIFGLTEGWRRFQDQLRPHHPLISTTAWHSLKLTHTDFTNIVNITPD 1596  
 QY 87 SAMPEGYVOERTIFFKD-----DGNKTRAE-----VKPEGDTLVNRIELKGDIFKEDG 135  
 Db 1597 SILPEALAAQOSVIAVQNTPPQSPYKREGECIIITDLPANGAIIINQQLKLPSTLKPSD 1656  
 QY 136 NILGHKLEYNVNS--HNVYIMADKQKNGIKVNFKIRHNIEDGSVOLADHYQOQNTPIGDGP 193  
 Db 1657 NI-----AEFPHQSIIKIIYI-ACQDNITECN-NILHLVQ--TLIKTQHY----- 1698  
 QY 194 VLLPDNHYLSTQSALS 209

Db 1699 ---PINLWLVTOGAIS 1711

RESULT 11

D71614

hypothetical protein PFB0460c - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C>Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000

C:Accession: D71614

R:Gardner, M.J.; Tetelkin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O. Science 282, 1128-1132, 1998

A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; MUID:99021743; PMID:9804551

C:Accession: D71614

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2573 <GAR>

A:Cross-references: GB:AE001362; NID:g3845188; PIDN:AAC71881.1; PID:g384519

A:Experimental source: clone 3D7

C:Genetics:

A:Gene: PFB0460c

Query Match 6.9%; Score 87.5; DB 2; Length 2573;

Best Local Similarity 26.2%; Pred. No. 1.2e+02;

Matches 34; Conservative 30; Mismatches 53; Indels 13; Gaps 5;

QY 94 VQERTIFFKD--DGNVKTAEVKFEGDTLVNRILKAGIDPKEDGNTLGHKLEYN--YNH 149

Db 126 LKKEITLKKD--DGNVKTAEVKFEGDTLVNRILKAGIDPKEDGNTLGHKLEYN--YNH 149

QY 150 NVYIMADKQNGIKVNFKIRHNIEDGSQLADHYQNTPTIGDGPVLLPDNHYLSTQSALS 209

Db 184 NLHIKENKKDEKKNKHNNDDNNM---IYYKNI---DKTHYLDNNVVHILNDIN 236

QY 210 KDPNEKRDHM 219

Db 237 TVLKRERYDM 246

RESULT 12

KXRTCI

proteinkin convertase 1 (EC 3.4.21.93) precursor - rat

N:Alternate names: furin homolog PC1; kexin homolog PC1; prohormone cleavage enzyme; pro

C:Species: Rattus norvegicus (Norway rat)

C>Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-May-2000

C:Accession: A41556; S27361; S36358

R:Bloomquist, B.T.; Eipper, B.A.; Mains, R.E. Mol. Endocrinol. 5, 2014-2024, 1991

A:Title: Prohormone-converting enzymes: regulation and evaluation of function using anti

A:Reference number: A41556; MUID:92168040; PMID:1791845

A:Accession: A41556

A:Molecule type: mRNA

A:Residues: 1-752 <BLO>

A:Cross-references: GB:M76705; NID:g203508; PIDN:AAA40945.1; PID:g203509

R:Hakes, D.J.; Birch, N.P.; Mezey, A.; Dixon, J.E. Endocrinology 129, 3053-3063, 1991

A:Title: Isolation of two complementary deoxyribonucleic acid clones from a rat insulinl

to endocrine and neuroendocrine tissues in rats.

A:Reference number: S27361; MUID:92063860; PMID:1954888

A:Accession: S27361

A:Molecule type: mRNA

A:Residues: 1-513,'A',515-752 <HAK>

A:Cross-references: EMBL:M83745

R:Hakes, D.J.; Birch, N.P.; Mezey, A.; Dixon, J.E. submitted to the EMBL Data Library, February 1992

A:Reference number: S36358

A:Accession: S36358

A:Molecule type: mRNA

A:Residues: 1-366,'T',367-513,'A',515-752 <HA2>

A:Cross-references: EMBL:M83745; NID:g205062; PIDN:AAA41476.1; PID:g205063

C:Comment: This protein is a member of a family of subtilisin-like proteinases responsib

C:Comment: This protein lacks a classical hydrophobic transmembrane segment but may asss

C:Superfamily: kexin; subtilisin homology

F:Keywords: glycoprotein; hydrolase; serine proteinase

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-110/Domain: propeptide #status predicted <PRO>

F:111-752/Product: prohormone-processing proteinase PC1 #status predicted <MAT>

F:158-396/Domain: subtilisin homology <SBT>

F:158-396/Domain: subtilisin homology <SBT>

F:173,401.645/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 6.8%; Score 87; DB 1; Length 752;

Best Local Similarity 25.6%; Pred. No. 28;

Matches 53; Conservative 20; Mismatches 70; Indels 64; Gaps 10;

QY 18 ELDGDVNGHK----FSVSGEGEGDATYGLTKLTKFICTTGKLP-----VPMPTLVLTGLSYG 68

Db 540 ERDTSFNGFKNDVMSVHTWGENPV--GTWTLKVTDMSGRMQMGEGRIVNWKLILHGT--- 594

QY 69 VOCFSRYPDHMKQHDFFKSKAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLVNRILKLG 128

Db 595 ----SSQPEHMKQ-----PRVYT-----SYNTVQNDRRGVKMKVNVVVEEKP 631

QY 129 IDPKEDGNILGHKLEYNVNSHNYIMADKQNGIKVNFKIRHNIEDGSQLADHYQNTPT 188

Db 632 TQNSLNGNLVPK---NSSSSVEDRRDEQVQAPSKAMLR-----LLQSAFSAKNTPT 680

QY 189 IGDGPVLLPDNHYLSTQSALSADPKNEK 215

Db 681 -----SKQS--SKIPSAK 691

RESULT 13

E82590

leucyl-tRNA synthetase XF2176 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000

R:Anonymous; The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

C:Accession: E82590

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717; PMID:10910347

A>Note: for a complete list of authors see reference number A59328 below

A:Accession: E82590

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-887 <SIM>

A:Cross-references: GB:AE004031; GB:AE003849; NID:g9107309; PIDN:AAF84975.1; GSPDB:GN0013

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; AJ

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme

J.D.; Junqueira, M.B.; Kemper, E.L.; Kitajima, J.P.; Krueger, J.E.; Kuramae, E.E.; Laigret

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaka

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira

M.; Tshunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF2176

C:Superfamily: leucine-tRNA ligase

Query Match 6.8%; Score 86.5; DB 2; Length 887;

Best Local Similarity 22.2%; Pred. No. 38;

Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;

QY 50 TTGKLPVPMPTLVLTGLSYGVQCFSRYPDHMKQHDFFKSKAMPEGYVQERTIFFKDDGNY-- 107

Db 329 TNEQLPV-VWANFVLMAYGTGAVMVFQHDQRDEQF--ANKYGLPIRQVIALKEPKNQDE 385

```

F;5-77/Domain: lipoyl/biotin-binding homology <LPB>
F;117-145/Region: beta-alpha-beta FAD nucleotide-binding fold
F;119-561/Domain: dihydrolipoamide dehydrogenase homology <LBD>
F;287-315/Region: beta-alpha-beta NAD nucleotide-binding fold
F;153-158/Disulfide bonds: redox-active #status predicted

Query Match          6.7%; Score 86; DB 1; Length 578;
Best Local Similarity 23.5%; Pred. No. 24;
Matches 54; Conservative 39; Mismatches 85; Indels 52; Gaps 12;

QY      10  TGVVPTLVLDGVNGHKFSVSGEGGDATYKGLTK-----FICTTGKLPVWPPTLVT 63
         |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db      255  TGSMPPIPEIE-----GNKLS-----GVIDST-GALSLESNPESIALIGGGVIGVEFASIFN 305

QY      64  GLSYGVQCFSRYPDHMKQHDFFKSAMPEGVQBERTIFFKDDGNKYKTAEVKFEQDILVNR 123
         |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db      306  SLGCKVSVIIEMLPHPILPMDREITSEI-----AKAKLIRDGININNN 346

QY      124  IELKGDIFKEDG---NILGHLEVYVNSHNVIYMADKOKN--GIKVNFKIRHNIEDGSVQ 178
         |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db      347  CKVTRIEQGEDGLKVFYIGDKGESIDVEKVLAVGRSNIIEGLDVE-KIGVKTGGSI 405
         |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:

QY      179  LADHYQONT-----PIGD--GPVLPLPNHYLSTQSALSKD-----PNEKRHD 218
         |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db      406  VNDKMETNVEGIYAIQDCTCKIMLA--HVASDQGVAAENIMGQNKMDY 453
         |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:

Search completed: June 21, 2004, 16:01:54
Job time : 12.3333 secs

```

**RESULT 15**

I40794  
dihydrolipoamide dehydrogenase (EC 1.8.1.4) [validated] - Clostridium magnum  
N;Alternate names: 2-oxoglutarate dehydrogenase complex chain E3; acetoin dehydrogenase  
dehydrogenase complex chain E3; S-complex 50K chain  
C;Species: Clostridium magnum  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
R;Accession: I40794  
R;Kruger, N.; Oppermann, F.B.; Lorenzl, H.; Steinbuchel, A.  
J. Bacteriol. 176, 3614-3630, 1994  
A;Title: Biochemical and molecular characterization of the Clostridium magnum acetoind  
A;Reference number: I40789; MUID:94266715; PMID:8206840  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-578 <KRU>  
A;Cross-references: GB:L31844; NID:g472324; PIDN:AAA21748.1; PID:g472330  
C;Function:  
A;Description: catalyzes the oxidation of dihydrolipoamide to lipoamide using NAD  
A;Pathway: acetoind dehydrogenase enzyme system  
C;Superfamily: Alkaligenes dihydrolipoamide dehydrogenase; dihydrolipoamide dehydrogenas  
e reductases; nnp\_oxidoreductae; redox-active disulfide



[7]  
X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.  
MEDLINE=98455509; PubMed=9782051;  
Wachter R.M., Elsiger M.A., Kallio K., Hanson G.T., Remington S.J.,  
"Structural basis of spectral shifts in the yellow-emission variants  
of green fluorescent protein.";  
Structure 6:1267-1277(1998).  
[8]  
X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
MEDLINE=99238303; PubMed=10220315;  
Elsiger M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J.,  
"Structural and spectral response of green fluorescent protein  
variants to changes in pH.";  
Biochemistry 38:5296-5301(1999).  
-1- FUNCTION: Energy-transfer acceptor. Its role is to transduce the  
blue chemiluminescence of the protein aequorin into green  
fluorescent light by energy transfer. Fluoresces in vivo upon  
receiving energy from the Ca(2+)-activated photoprotein aequorin.  
Absorbs light maximally at 395 nm and exhibits a smaller  
absorbance peak at 470 nm. The fluorescence emission spectrum  
peaks at 509 nm with a shoulder at 540 nm.  
-1- SUBUNIT: Monomer.  
-1- TISSUE SPECIFICITY: Photocytes.  
-1- PFM: Contains a covalently attached chromophore, which is composed  
of modified amino acid residues. The chromophore is formed upon  
cyclization of the residues Ser-dehydrotyr-Gly.  
-1- BIOTECHNOLOGY: Has become a useful and ubiquitous tool for making  
chimeric proteins of GFP linked to other proteins where it  
functions as a fluorescent protein tag. GFP tolerates N- and C-  
terminal fusion to a broad variety of proteins. It has been  
expressed in bacteria, yeast, slime mold, plants, Drosophila,  
zebrafish, and in mammalian cells. As a noninvasive fluorescent  
marker in living cells, it allows for a wide range of applications  
where it may function as a cell lineage tracer, reporter of gene  
expression, or as a measure of protein-protein interactions.  
-1- DATABASE: NAME-Protein Spotlight;  
NOTE=Issue 11 of June 2001;  
WWW="http://www.expasy.org/spotlight/articles/sptlt011.html".  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M62654; AAA27722.1; -;  
EMBL; M62653; AAA27721.1; -;  
EMBL; L29345; AAA58246.1; -;  
EMBL; X96418; CAA65278.1; -;  
PIR; JS0692; JQ1514.  
PDB; 1B9C; 17-NOV-00.  
PDB; 1BFP; 07-JUL-97.  
PDB; 1CAF; 14-JUN-00.  
PDB; 1EMA; 08-NOV-96.  
PDB; 1EMB; 16-JUN-97.  
PDB; 1EMC; 20-AUG-97.  
PDB; 1EME; 20-AUG-97.  
PDB; 1EMF; 20-AUG-97.  
PDB; 1EMG; 12-MAY-99.  
PDB; 1EMK; 20-AUG-97.  
PDB; 1EML; 20-AUG-97.  
PDB; 1EMM; 20-AUG-97.  
PDB; 1F09; 17-NOV-00.  
PDB; 1FOB; 17-NOV-00.  
PDB; 1GFL; 11-JAN-97.  
PDB; 1HCJ; 15-JAN-02.  
PDB; 1HUY; 04-JUL-01.  
PDB; 1JBY; 07-JAN-03.  
PDB; 1JBZ; 07-JAN-03.  
PDB; 1KP5; 28-AUG-02.  
PDB; 1KYP; 10-APR-02.

DR PDB; 1KYR; 10-APR-02.  
DR PDB; 1KYS; 10-APR-02.  
DR PDB; 1YFP; 28-OCT-98.  
DR PDB; 2EMD; 20-AUG-97.  
DR PDB; 2EMN; 20-AUG-97.  
DR PDB; 2EMO; 20-AUG-97.  
DR InterPro; IPR009017; GFP like.  
DR InterPro; IPR000786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFLUORESCENT.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
KW Luminescence; 3D-structure.  
FT CROSSLINK 65 67 5-imidazolinone (Ser-Gly).  
FT MOD RES 66 66 2,3-DIDEHYDROTYROSINE.  
FT VARIANT 100 100 F -> Y.  
FT VARIANT 108 108 T -> S.  
FT VARIANT 141 141 L -> M.  
FT VARIANT 219 219 V -> I.  
FT CONFLICT 2 2 S -> G (IN REF. 3).  
FT CONFLICT 25 25 H -> Q (IN REF. 2).  
FT CONFLICT 80 80 Q -> R (IN REF. 3).  
FT CONFLICT 157 157 Q -> P (IN REF. 2).  
FT CONFLICT 172 172 E -> K (IN REF. 2).  
FT HELIX 4 8  
FT STRAND 12 22  
FT TURN 23 24  
FT STRAND 25 36  
FT TURN 37 40  
FT STRAND 41 48  
FT TURN 49 50  
FT TURN 57 60  
FT TURN 61 63  
FT HELIX 69 71  
FT STRAND 73 73  
FT HELIX 76 81  
FT HELIX 83 86  
FT TURN 87 90  
FT STRAND 92 100  
FT TURN 101 102  
FT STRAND 105 115  
FT TURN 116 117  
FT STRAND 118 128  
FT TURN 132 133  
FT TURN 135 139  
FT STRAND 141 141  
FT STRAND 148 155  
FT TURN 156 159  
FT STRAND 160 171  
FT TURN 172 173  
FT STRAND 176 187  
FT STRAND 199 208  
FT TURN 211 212  
FT STRAND 217 227  
SQ SEQUENCE 238 AA; 26886 MW; EA5A6F21FBFB6E05 CRC64;  
Query Match 97.3%; Score 1240; DB 1; Length 238;  
Best Local Similarity 97.9%; Pred. No. 1.3e-95;  
Matches 233; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
Qy 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDGYGKLTFLKFICTTGKLPVWPPTL 61  
Db 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDGYGKLTFLKFICTTGKLPVWPPTL 60  
Qy 62 VTGLSYGVQCFSRYPDHMKQHDFFKSAAPGEGYVQERTIPFKDGNKYNTRAEVKEGDTLV 121  
Db 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAAPGEGYVQERTIPFKDGNKYNTRAEVKEGDTLV 120  
Qy 122 NRTELKIDFKEDGNLTGLHKLKLVNYSNHYINADKQKNGIKVNFKIRHNIEDGSVOLAD 181  
Db 121 NRTELKIDFKEDGNLTGLHKLKLVNYSNHYINADKQKNGIKVNFKIRHNIEDGSVOLAD 180  
Qy 182 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNKRDRHMLLGFVTAAGITLGMDELYK 239  
Db 182 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNKRDRHMLLGFVTAAGITLGMDELYK 239

Db	181	HYQNTPIGDGPVLLPDNHYLSTQSLKSDPNKRDHNVLLFEVTAAGITHGMDELK 238	
<p>RESULT 2</p> <p>VIT4 CAEEL STANDARD; PRT; 1603 AA.</p> <p>AC P18947; Q9BPP3; 01-NOV-1990 (Rel. 16, Created)</p> <p>DT 01-OCT-2003 (Rel. 42, Last sequence update)</p> <p>DT 10-OCT-2003 (Rel. 42, Last annotation update)</p> <p>DE Vitellogenin 4 precursor.</p> <p>GN VIT-4 OR F59D8.2.</p> <p>OS Caenorhabditis elegans.</p> <p>OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;</p> <p>OC Rhabditidae; Peloderinae; Caenorhabditis.</p> <p>OX NCBI_TaxID=6239;</p> <p>LN [1]</p> <p>RP SEQUENCE FROM N.A.</p> <p>RC STRAIN=Bristol N2;</p> <p>RA Waterston R.;</p> <p>RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.</p> <p>LN [2]</p> <p>RP SEQUENCE OF 1-282 FROM N.A.</p> <p>RA Blumenthal T., Spieth J., Zucker E.;</p> <p>RL Submitted (NOV-1986) to the EMBL/GenBank/DBJ databases.</p> <p>LN [3]</p> <p>RP SEQUENCE OF 1-71 FROM N.A.</p> <p>RA MEDLINE=85269643; PubMed=4022780;</p> <p>RX Spieth J., Denison K., Kirtland S., Cane J., Blumenthal T.;</p> <p>RT "The C. elegans vitellogenin genes: short sequence repeats in the promoter regions and homology to the vertebrate genes.";</p> <p>RL Nucleic Acids Res. 13:5283-5295(1985).</p> <p>CC -!- FUNCTION: Precursor of the egg-yolk proteins that are sources of nutrients during embryonic development (Potential).</p> <p>CC -!- SUBCELLULAR LOCATION: Secreted.</p> <p>CC -!- TISSUE SPECIFICITY: Synthesized in Caenorhabditis only by 32 cells building the intestine of adult hermaphroditic individuals; they are cotranslationally secreted into the body cavity and subsequently taken up by the gonad.</p> <p>CC -!- SIMILARITY: Contains 1 WVPD domain.</p> <p>CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announcement/">http://www.isb-sib.ch/announcement/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a>).</p> <p>CC EMBL; AC024137; AAK09074.1; -.</p> <p>DR EMBL; M11498; AAA28163.1; -.</p> <p>DR EMBL; X02754; CAA26531.1; -.</p> <p>DR PIR; A43084; A43084.</p> <p>DR WormPep; F59D8.2; CE26817.</p> <p>DR InterPro; IPR001747; Lipid transprt_N.</p> <p>DR InterPro; IPR001846; WVF D.</p> <p>DR Pfam; PF01347; Vitellogenin_N; 1.</p> <p>DR SMART; SM00638; LPD_N; 1.</p> <p>DR SMART; SM00216; VWD; 1.</p> <p>DR Storage protein; Multigene family; Signal.</p> <p>KW SIGNAL 1 15 POTENTIAL.</p> <p>FT CHAIN 16 1603 VITELLOGENIN 4.</p> <p>FT DOMAIN 1308 1455 VWF.</p> <p>FT FT CONFLICT 30 30 Y -&gt; V (IN REF. 3).</p> <p>FT FT CONFLICT 169 169 L -&gt; V (IN REF. 2).</p> <p>FT FT CONFLICT 183 187 EVAYT -&gt; RSRLLH (IN REF. 2).</p> <p>FT FT CONFLICT 275 275 T -&gt; S (IN REF. 2).</p> <p>SQ SEQUENCE 1603 AA; 186307 MW; E303170325BC99B CRC64;</p> <p>Query Match 7.4%; Score 94.5; DB 1; Length 1603;</p> <p>Best Local Similarity 23.9%; Pred. No. 7.8;</p> <p>Matches 53; Conservative 32; Mismatches 68; Indels 69; Gaps 12;</p>			
Qy	1	MVSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDGYTKLTKFTCTTGKLPVPMPT 60	
Db	162	MESDKDSLFFNVHEKTMGDCV---AYTIQVQG-GKTIYTKSVNFDKCIIR-----PE 211	
Qy	61	LVTGLSYGVQCFSPRYDPHMKQHDPPKSAWPEG-VYQERTIF---FKDDG----- 105	
Db	212	TAYGLRFGSEC-----KCEKEGQFVQPTVYTFYFQEKLOESEVNSIYT 257	
Qy	106	-----NYKTRAEVKPEGDTLVNRIELKIDFKEDGNILGHKLEYNYNHNVYINAD 156	
Db	258	LVNNGQEVVVKSETRAKTVFVEESKINR-EIK-----KVSQPKKEIIVYSMENEKLEIQ 308	
Qy	157	KQKNG-----IKVNFKIRHNIEDSGVQLADHYQOQNTP 188	
Db	309	FYKQGDKAQVNPFAIEIKQV-EQLEEIPFRQIOEH-EQNTP 348	
<p>RESULT 3</p> <p>D152_HAEIN STANDARD; PRT; 795 AA.</p> <p>AC P44935; 01-NOV-1995 (Rel. 32, Created)</p> <p>DT 01-NOV-1995 (Rel. 32, Last sequence update)</p> <p>DT 28-FEB-2003 (Rel. 41, Last annotation update)</p> <p>DE Protective surface antigen D15 precursor (80 kDa D15 antigen)</p> <p>DE (D-15-Ag) (Outer membrane protein D15).</p> <p>GN HI0917.</p> <p>OS Haemophilus influenzae.</p> <p>OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;</p> <p>OC Pasteurellaceae; Haemophilus.</p> <p>OX NCBI_TaxID=727;</p> <p>LN [1]</p> <p>RP SEQUENCE FROM N.A.</p> <p>RC STRAIN=Rd / KW20 / ATCC 51907;</p> <p>RX MEDLINE=95350630; PubMed=7542800;</p> <p>RA Kleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kersavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M., Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;</p> <p>RA "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";</p> <p>RT Science 269:496-512(1995).</p> <p>RL CC -!- SUBCELLULAR LOCATION: Outer membrane.</p> <p>CC -!- SIMILARITY: Belongs to the surface antigen D15 family.</p> <p>CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announcement/">http://www.isb-sib.ch/announcement/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a>).</p> <p>CC EMBL; U32773; AAC22575.1; -.</p> <p>DR TIGR; HI0917; -.</p> <p>DR InterPro; IPR000184; Bac_surfAg_D15.</p> <p>DR Pfam; PF01103; Bac surface Ag; 1.</p> <p>KW Antigen; Outer membrane; Signal; Complete proteome.</p> <p>FT SIGNAL 1 19 POTENTIAL.</p> <p>FT CHAIN 20 795 PROTECTIVE SURFACE ANTIGEN D15.</p> <p>SQ SEQUENCE 795 AA; 87478 MW; B85691FC22E6ED44 CRC64;</p> <p>Query Match 7.3%; Score 92.5; DB 1; Length 795;</p> <p>Best Local Similarity 22.3%; Pred. No. 4.9;</p> <p>Matches 49; Conservative 30; Mismatches 78; Indels 63; Gaps 11;</p>			
Qy	64	GLSYGVQCFSPRYDPHMKQHD-----FKSAMPEGYVQES-----RT 98	

```
Db 426 GIGYGTEGSGISQASVKQDNFLGTGAASVIACTKNDYGTSVNLGYTEPYFTTKDGVSLGNN 485
QY 99 IFPKDGNKYKTRAEVKFEGDGLVNRLELKGIDFKEDGNI---LGH-----KLEYNYN 147
Db 486 VFPENYDNSKSDTSNRYKRTTYGNSVTL-GFPVNNNSYYVGLGHTYNNKISNFALEYN-- 542
QY 148 SHNVYIMADKQK-NGIKVNFKIRHNIEDGVSQVLADHYQQ-----NTPIGDGPVL 195
Db 543 -RNLYIQSMKFKNGIKTN-----DFDFSGWYNSLNRGYFPFTKGVKASLG-GRVT 592
QY 196 LP--DNHLYSTQSALS KDPNEKRDHMLVLLGFTVTAAGITLG 233
Db 593 IFGSDNKKYKLSADVQGFYPLDRDLHLWVVSASAKASAGYANG 632
```

## RESULT 4

```
ID D151 HAEIN STANDARD; PRT; 797 AA.
AC P46024;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype B;
RX MEDLINE=95255676; PubMed=7737523;
RA Flack F.S., Loosmore S., Chong P., Thomas W.R.;
RT "The sequencing of the 80-kDa D15 protective surface antigen of
RT Haemophilus influenzae."
RL Gene 156:97-99(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Minna / Serotype B, and Eagan / Serotype B;
RX MEDLINE=97427952; PubMed=9284140;
RA Loosmore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease."
RL Infect. Immun. 65:3701-3707(1997).
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: Belongs to the surface antigen D15 family.
```

```
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U13961; AAA85645.1; -
DR EMBL; U60832; AAB61974.1; -
DR EMBL; U60833; AAB61976.1; -
DR PIR; JC4078; JC4078.
DR InterPro; IPR000184; Bac_surfAg_D15.
DR Pfam; PF01103; Bac_surface_Ag_1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 19
FT CHAIN 20 797 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 797 AA; 87675 MW; 2F93DE538696AF18 CRC64;
```

Query Match 7.3%; Score 92.5; DB 1; Length 797;

Best Local Similarity 22.3%; Pred. No. 4.9; Mismatches 63; Gaps 11;

Matches 49; Conservative 30; Indels 78; Indels 63; Gaps 11;

64 GLSYGVQCFGRYPDHMKQHDF-----FKSAMPEGYVQE-----RT 98

```
Db 426 GIGYGTEGSGISQASVKQDNFLGTGAASVIACTKNDYGTSVNLGYTEPYFTTKDGVSLGNN 485
QY 99 IFPKDGNKYKTRAEVKFEGDGLVNRLELKGIDFKEDGNI---LGH-----KLEYNYN 147
Db 486 VFPENYDNSKSDTSNRYKRTTYGNSVTL-GFPVNNNSYYVGLGHTYNNKISNFALEYN-- 542
QY 148 SHNVYIMADKQK-NGIKVNFKIRHNIEDGVSQVLADHYQQ-----NTPIGDGPVL 195
Db 543 -RNLYIQSMKFKNGIKTN-----DFDFSGWYNSLNRGYFPFTKGVKASLG-GRVT 592
QY 196 LP--DNHLYSTQSALS KDPNEKRDHMLVLLGFTVTAAGITLG 233
Db 593 IFGSDNKKYKLSADVQGFYPLDRDLHLWVVSASAKASAGYANG 632
```

## RESULT 5

```
GUN_BACS6
ID GUN_BACS6 STANDARD; PRT; 941 AA.
AC P19424;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Alkaline cellulase).
OS Bacillus sp. (strain KSM-635).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1415;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91037937; PubMed=2230718;
RA Ozaki K., Shikata S., Kawai S., Ito S., Okamoto K.;
RT "Molecular cloning and nucleotide sequence of a gene for alkaline
RT cellulase from Bacillus sp. KSM-635."
RL J. Gen. Microbiol. 136:1327-1334(1990).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -1- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC -1- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
```

```
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M27420; AAA22304.1; -
DR PIR; S29043; S29043.
DR PDB; 1G01; 31-DEC-02.
DR PDB; 1G0C; 31-DEC-02.
DR InterPro; IPR005086; CBM_17_28.
DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR001547; Glyco_hydro_5.
DR InterPro; IPR001119; SLH.
DR Pfam; PF03424; CBM_17_28; 2.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00395; SLH; 3.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR PROSITE; PS01072; SLH_DOMAIN; 2.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat;
KW 3D-structure.
FT SIGNAL 1 29
FT CHAIN 30 941 ENDOGLUCANASE.
FT DOMAIN 40 99 SLH 1.
FT DOMAIN 100 151 SLH 2.
FT DOMAIN 152 225 SLH 3.
FT ACT_SITE 373 373 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 485 485 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 941 AA; 104628 MW; BEA2AC3B169BFADA CRC64;
```



```
Query Match          7.3%; Score 92.5; DB 1; Length 941;
Best Local Similarity 20.2%; Pred. No. 6;
Matches 49; Conservative 34; Mismatches 60; Indels 99; Gaps 11;

QY 16 LVELDGVNGHKFSVSGEGSDATYGKLTCLKFICTTGKLPVWPVTLVTGLS-YGVQCFSR 74
DB 240 LVSENG-----QUTLAGE---DGT-----PVQLRGMSHTGLQWFG- 271

QY 75 YPDHMKQHDFFKSGAMPEGYVQERTIFFKDDGNKYKTRAEVKEGDTLVNRIELKGIDFKED 134
DB 272 --BIVNENAFVSLNDWGSNWIRLAWYIGENGATNPEVK--DLVYEGIELA----- 319

QY 135 GNTLGHKLEYNHNVNHYIMADKQKGIKVNFKIRNIEDGSV-----QLADHYQOQNTPIG 190
DB 320 -----FEHDMYIVTDWH---VHAPGDPRADVYSGAYDFFEETADHYKDH---- 360

QY 191 DGPVLVLPDHYLSTQSALSXDPN-----EKRDMHVL 221
DB 361 -----PKNHYIILWELANSPNNGGPGLTNDEKGEAVKEVAPIVEMLRKGDNMIL 414

QY 222 LG 223
DB 415 VG 416

RESULT 6
CP51_CANGA
ID CP51_CANGA STANDARD; PRT; 533 AA.
AC P50859; Q02312;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 51 (EC 1.14.13.70) (CYP51) (P450-LIA1) (Sterol 14-
DE alpha-demethylase) (Lanosterol 14-alpha demethylase) (P450-14DM).
GN ERG11 OR CYP51.
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2001-L5;
RX MEDLINE=96161286; PubMed=8593007;
RA Geber A., Hitchcock C.A., Swartz J.E., Pullen F.S., Marsden K.E.,
RA Kwon-Chung K.J., Bennett J.E.;
RT "Deletion of the Candida glabrata ERG3 and ERG11 genes: effect on cell
RT viability, cell growth, sterol composition, and antifungal
RT susceptibility.";
RL Antimicrob. Agents Chemother. 39:2708-2717(1995).
RN [2]
RP SEQUENCE OF 60-473 FROM N.A.
RC STRAIN=ATCC 2001;
RX MEDLINE=95081364; PubMed=7989540;
RA Burgener-Kairuz P., Zuber J.P., Jaunin P., Buchman T.G., Bille J.,
RA Rossier M.;
RT "Rapid detection and identification of Candida albicans and
RT Torulopsis (Candida) glabrata in clinical specimens by
RT species-specific nested PCR amplification of a cytochrome P-450
RT lanosterol-alpha-demethylase (LIA1) gene fragment.";
RL J. Clin. Microbiol. 32:1902-1907(1994).
CC -!- FUNCTION: Catalyzes C14-demethylation of lanosterol which is
CC critical for ergosterol biosynthesis. It transforms lanosterol
CC into 4,4'-dimethyl cholesta-8,14,24-triene-3-beta-ol (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: Obtusifoliol + 3 O(2) + 3 NADPH = 4-alpha-
CC methyl-5-alpha-ergosta-8,14,24(28)-trien-3-beta-ol + formate + 3
CC NADP(+) + 3 H(2)O.
CC -!- PATHWAY: Ergosterol biosynthesis.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR EMBL: J50834; AAB61977.1; -
DR InterPro; IPR000184; Bac_surfAg_D15.
DR Pfam; PF01103; Bac surface Ag; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 193 POTENTIAL.
FT CHAIN 20 793 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 793 AA; 87511 MW; 51FDB2036801A14 CRC64;

Query Match
Best Local Similarity 7.2%; Score 91.5; DB 1; Length 793;
Matches 50; Conservative 28; Mismatches 79; Indels 63; Gaps 11;

QY 34 GLSYGVQCFRYPDHMKQHDFF-----FKSAMPEGVYQEE-----RT 98
DB 426 GIGYGTESGYSQTSIKQDNFLGTGAASVAGTKNDYGTGVSNLGYTEPYFTKDGVSLOGN 485
QY 99 IFPKDGNKVTAEVFEFGDTLVNRTELKGDIDPKDGNI---LGH-----KLENNYN 147
DB 496 IFPENYDNSKDTSSNYKRTYGSNTVL-GFPVNNNSYVGLGHTYKISNPALEYN-- 542
QY 148 SHNVYIMADQK-NGIKVNFKIRHNTEDSGVQLADHYQ-----NTPIGDGPVL 195
DB 543 -RNLYTQSMKFKNGIKTN-----DFDFSEGMWNSLNRGYFTPTKGKASLG-GRVT 592
QY 196 LP--DNHYSLTQSALSQDNEKEDHNVLLGLFVTAAGITLG 233
DB 593 IPGSDNKYYKLSADVQGVPLDRDHRVWSAKASGYANG 632

RESULT 8
SYL_HAEIN
ID SYL_HAEIN STANDARD; PRT; 861 AA.
AC P43827;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine-tRNA ligase) (LeuRS).
GN LEUS OR HI0921.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA McLaughlin A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA Kienley K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RD.";
RL Science 269:496-512(1995).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/ebis/announce/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL: U32774; AAC22581.1; -
```

```
DR PIR; H64102; H64102.
DR TIGR; HI0921; -.
DR HAMAP; MF_00049; -; 1.
DR InterPro; IPR002302; Leu-trNASynt1a.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR009008; ValRS_IleRS_edit.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRPFAM; TIGR00396; leus_bact; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_1; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
SQ SEQUENCE 861 AA; 97750 MW; EB93304F6B4C8FB7 CRC64;

Query Match
Best Local Similarity 6.9%; Score 88.5; DB 1; Length 861;
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;

QY 50 TTGKLPVPWPTLVGLSYGVQCFRYPDHMKQHDFFKSAPEGYVQERTIFFPKD----- 103
DB 314 TGDKLPI-WVANFVLMHYGTGAVMAVPAH-DQRF-----EFAQKYSLPKQVIAPLA 364
QY 104 DGNKTRAEVKPEGDTLVNRTELKGDIDPKEDGNILGHKLEYNNSHNVYIMADK-QKNGI 162
DB 365 DEEIDLTKQAFVEHGLKLVNSDFDGKPF--DGAFNG-----IADKLEKLG 408
QY 163 ---KVNFKIRH-----NIEDSGVQLADHYQQNTPIGDGVLPLPDHNYL- 202
DB 409 GKQVNVRLRDWGSVRSQRYGAPIMPLTLENGDVVPA-----PMEDLPILPEDVVD 461
QY 203 STQSALSQDNP 213
DB 462 GVKSPINADPN 472

RESULT 9
NECI_RAT
ID NECI_RAT STANDARD; PRT; 752 AA.
AC P28840;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neuroendocrine convertase 1 precursor (EC 3.4.21.93) (PC1)
DE (Prohormone convertase 1) (proprotein convertase 1).
GN PCSK1 OR NEC1 OR NEC-1 OR BDP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92168040; PubMed=1791845;
RA Bloomquist B.T., Eipper B.A., Mains R.E.;
RT "Prohormone-converting enzymes: regulation and evaluation of function
RD using antisense RNA.";
RL Mol. Endocrinol. 5:2014-2024 (1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92063860; PubMed=1954888;
RA Hakes D.J., Birch N.P., Mezey A., Dixon J.E.;
RT "Isolation of two complementary deoxyribonucleic acid clones from a
RT rat insulinoma cell line based on similarities to Kex2 and furin
RT sequences and the specific localization of each transcript to
RT endocrine and neuroendocrine tissues in rats.";
RL Endocrinology 129:3053-3063(1991).
CC -1- FUNCTION: Involved in the processing of hormone and other protein
CC precursors at sites comprised of pairs of basic amino acid
CC residues. Substrates include POMC, renin, enkephalin, dynorphin,
CC somatostatin and insulin.
CC
```

CC -1- CATALYTIC ACTIVITY: Release of protein hormones, neuropeptides and  
CC bonds from their precursors, generally by cleavage of -Lys-Arg-|-  
CC  
CC  
CC -1- COFACTOR: Calcium.  
CC  
CC -1- SUBCELLULAR LOCATION: Localized in the secretion granules.  
CC  
CC -1- SIMILARITY: Belongs to peptidase family S8. Furin subfamily.  
CC  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M76705; AAA40945.1; -;  
CC EMBL; M83745; AAA41476.1; -;  
CC PIR; A43556; KRTCL.  
CC HSP; Q45670; IDBI.  
CC MEROPS; S08.072; -;  
CC InterPro; IPR000209; Peptidase S8.  
CC InterPro; IPR002884; Peptidase S8B.  
CC InterPro; IPR009020; Protease Inhib.  
CC Pfam; PF01483; P\_protein; 1.  
CC Pfam; PF00082; Peptidase S8; 1.  
CC PRINTS; PR00723; SUBTILISIN.  
CC ProDom; PD00717; P domain; 1.  
CC PROSITE; PS00136; SUBTILASE ASP; 1.  
CC PROSITE; PS00137; SUBTILASE HIS; 1.  
CC PROSITE; PS00138; SUBTILASE SER; 1.  
CC Hydrolase; Serine protease; Glycoprotein; Zymogen; Calcium; Signal.  
FT SIGNAL 1 27  
FT CHAIN 28 110 POTENTIAL.  
FT CHAIN 111 752 NEUROENDOCRINE CONVERTASE 1.  
FT DOMAIN 122 410 CATALYTIC.  
FT DOMAIN 739 751 AMPHIPATHIC (POTENTIAL).  
FT ACT\_SITE 167 167 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 208 208 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 382 382 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT CARBOHYD 173 173 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 366 366 T -> TT (IN REF. 2).  
FT CONFLICT 514 514 E -> A (IN REF. 2).  
SQ SEQUENCE 752 AA; 84120 MW; F630AD830A076DED CRC64;  
  
Query Match 6.8%; Score 87; DB 1; Length 752;  
Best Local Similarity 25.6%; Pred. No. 13;  
Matches 53; Conservative 20; Mismatches 70; Indels 64; Gaps 10;  
  
Qy 18 ELDDGVNGHK- - - - -RSVSGEGEGDATYGLTLKFICTGKLP- - - - -VPWPTLVTLGSLYG 68  
Db 540 ERDTSNGFANDFMSVHTWGENPV- - -GTWTLKVTMSGRMQNEGRVNWKLILHGT- - - - -594  
  
Qy 69 VQCFSRYPDMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKG 128  
Db 595 - - - - -SSQPEHMQ- - - - -PRVYT- - - - -SYNTQNDRRGVEKMNVVVEKP 631  
  
Qy 129 IFKEDCNILGHKLEYNVNSHNVYIMADKQKGIKVKFKIRNIHEDGSGVOLADHYQONTTP 188  
Db 632 TQNSLGNLLVLPK- - - - -NSSSSSVEDRRDEQVQGAPSKAMLR- - - - -LLQSAFSKNTP 680  
  
Qy 189 IGDGPVLLPDNHYLSTQSLSKDPNEK 215  
Db 681 - - - - -SKQS- - - - -SKIPSAK 691  
  
RESULT 10  
ID SYL\_XYLEA STANDARD; PRT; 879 AA.  
AC Q9PBG8;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LEURS).  
GN LEUS OR XF2176.  
OS Xylella fastidiosa.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
OC Xanthomonadaceae; Xylella.  
OX NCBI\_TaxID=2371;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=9A5C;  
RX MEDLINE=20365717; PubMed=10910347;  
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
RA Alvarado R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,  
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,  
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
RA Machado M.A., Madeira A.M.B.N., Martins E.M.F., Matsukuma A.Y.,  
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
RA Quadrio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,  
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
RL "The genome sequence of the plant pathogen Xylella fastidiosa.";  
Nature 406:151-159(2000).  
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +  
CC diphosphate + L-leucyl-tRNA(Leu).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AB004031; AAF84975.1; ALT\_INIT.  
DR HAMAP; MF\_00049; -; 1.  
DR InterPro; IPR002302; Leu-tRNA synthetase.  
DR InterPro; IPR002300; tRNA-synt 1a.  
DR InterPro; IPR001412; tRNA-synt 1.  
DR InterPro; IPR009008; ValRS\_1leuS\_edit.  
DR Pfam; PF00133; tRNA-synt 1; 1.  
DR PRINTS; PR00985; TRNASYNTHLEU.  
DR TIGRFAMs; TIGR00396; leuS bact; 1.  
DR PROSITE; PS00178; AA tRNA LIGASE I; 1.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KW Complete proteome.  
FT SITE 45 55 "HIGH" REGION.  
FT SITE 637 641 "KMSKS" REGION.  
FT BINDING 640 640 ATP (BY SIMILARITY).  
SQ SEQUENCE 879 AA; 99796 MW; 9FDCB992092919E CRC64;  
  
Query Match 6.8%; Score 86.5; DB 1; Length 879;  
Best Local Similarity 22.2%; Pred. No. 17;  
Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;

[illegible]

```

RESULT 11
TRNA_CAMJE STANDARD; PRT; 357 AA.
ID TRNA_CAMJE
AC Q9PP92;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE tRNA (Uracil-5)-methyltransferase (EC 2.1.1.35) (tRNA(M-5-U54))-
DE methyltransferase (RUMT).
DE TRMA OR CJ0831C.
GN NCBI_TaxID=197;
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX
OX NCBI_TaxID=197;
RN [1]
SEQUENCE FROM N.A.
RC
RC STRAIN=NCCT 11168;
RX MEDLINE=20150912; PubMed=10698204;
RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrall B.G.;
RA "The genome sequence of the food-borne pathogen Campylobacter jejuni
RA reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
CC -1- FUNCTION: Catalyzes the formation of 5-methyl-uridine at position
CC 54 (M-5-U54) in all tRNA (By similarity).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
CC homocysteine + tRNA containing thymine.
CC -1- SIMILARITY: Belongs to the RNA M5U methyltransferase family. Trma
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/ammouse/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL139076; CAB73096.1; --
CC PIR; G81355; G81355.
CC HAMAP; MF_010111; 1.
CC InterPro; IPR000051; SAM_bind.
CC InterPro; IPR001566; TrmA.
CC PROSITE; PS01230; TrmA 1; 1.
CC PROSITE; PS01231; TrmA 2; FALSE NEG.
CC Transferrase; Methyltransferase; tRNA processing; Complete proteome.
CC DOMAIN 207 213
CC S-ADENOSYLMETHIONINE BINDING (BY
CC SIMILARITY).
CC ACT_SITE 315 315 BY SIMILARITY.
CC SEQUENCE 357 AA; 42276 MW; CEC5328347CEE497 CRC64;
DR
DR PIR; G81355; G81355.
DR HAMAP; MF_010111; 1.
DR InterPro; IPR000051; SAM_bind.
DR InterPro; IPR001566; TrmA.
DR PROSITE; PS01230; TrmA 1; 1.
DR PROSITE; PS01231; TrmA 2; FALSE NEG.
DR Transferrase; Methyltransferase; tRNA processing; Complete proteome.
DR DOMAIN 207 213
DR S-ADENOSYLMETHIONINE BINDING (BY
DR SIMILARITY).
FT ACT_SITE 315 315 BY SIMILARITY.
FT SEQUENCE 357 AA; 42276 MW; CEC5328347CEE497 CRC64;
SQ

```

```
Db 16 KSELEFT-----LRKETFKORLANWAVKINGMEFDQYDDDDNATYLVFGVEGDQVICSRLIE 71
QY 56 VPMPTLVTLGSLYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKF 115
Db 72 TKYPNMITG-----TFFPY-----FEKIDPEGKYIESRFFVDKARSKTILGNSY 117
QY 116 EBDTL-----VNRIELKGIKDFKEDG--NILGHKL-----EYNYNGH 149
Db 118 PVSTMEFLATVNYSKSGY----DGVYTVSHPMILTKRSGWKISIVGQGMSEKHERVY 173
QY 150 NVYIMADKQKNGIKVNFKIRHNIE 173
Db 174 LLFLPVDNESQDVLVR-RINHNOE 196

RESULT 13
ITIH3_MESAU STANDARD; PRT; 886 AA.
AC P97280;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Inter-alpha-trypsin inhibitor heavy chain H3 precursor (ITI heavy
DE chain H3) (Inter-alpha-inhibitor heavy chain 3) (HC3).
GN ITIH3.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97420688; PubMed=9276673;
RA Nakatani T., Suzuki Y., Yamamoto T., Sinohara H.;
RT "Molecular cloning and sequencing of cDNAs encoding three heavy-chain
RT precursors of the inter-alpha-trypsin inhibitor in Syrian hamster:
RT implications for the evolution of the inter-alpha-trypsin inhibitor
RT heavy chain family.";
RL J. Biochem. 122:71-82(1997).
RN [2]
RP SEQUENCE OF 31-50; 446-472 AND 504-523, AND SUBUNITS.
RC TISSUE=Plasma;
RX MEDLINE=97018241; PubMed=8864857;
RA Yamamoto T., Yamamoto K., Sinohara H.;
RT "Inter-alpha-trypsin inhibitor and its related proteins in Syrian
RT hamster urine and plasma.";
RL J. Biochem. 120:145-152(1996).
CC -!- FUNCTION: MAY ACT AS A CARRIER OF HYALURONAN IN SERUM OR AS A
CC BINDING PROTEIN BETWEEN HYALURONAN AND OTHER MATRIX PROTEIN,
CC INCLUDING THOSE ON CELL SURFACES IN TISSUES TO REGULATE THE
CC LOCALIZATION, SYNTHESIS AND DEGRADATION OF HYALURONAN WHICH ARE
CC ESSENTIAL TO CELLS UNDERGOING BIOLOGICAL PROCESSES (BY
CC SIMILARITY).
CC -!- SUBUNIT: I-alpha-I plasma protease inhibitors are assembled from
CC one or two heavy chains (H1, H2 or H3) and one light chain,
CC bikunin. Inter-alpha-inhibitor (I-alpha-I) is composed of H1, H2
CC and bikunin, and inter-alpha-like inhibitor (I-alpha-Li) of H2 and
CC bikunin, and pre-alpha-inhibitor (P-alpha-I) of H3 and bikunin.
CC -!- PTM: Heavy chains are interlinked with bikunin via a chondroitin
CC 4-sulfate bridge to the their C-terminal aspartate (By
CC similarity).
CC -!- SIMILARITY: Belongs to the ITIH family.
CC -!- SIMILARITY: Contains 1 VWFA domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
DR EMBL; D89287; BAA13940.1; -.
DR InterPro; IPR006587; VIT.
DR InterPro; IPR002035; VWFA_A.
DR Pfam; PF00092; vwa; 1.
DR SMART; SM00609; VII; 1.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS0234; VWFA; 1.
KW Serine protease inhibitor; Repeat; Signal; Multigene family;
KW Glycoprotein.
FT SIGNAL 1 18 POTENTIAL.
FT PROPEP 19 30 BY SIMILARITY.
FT CHAIN 31 646 INTER-ALPHA-TRYPSIN INHIBITOR HEAVY CHAIN
FT H3.
FT PROPEP 647 886 BY SIMILARITY.
FT DOMAIN 279 439 VWFA.
FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 577 577 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT BINDING 646 646 CHONDROITIN 4-SULFATE, CROSS-LINK SITE
FT (BY SIMILARITY).
SQ SEQUENCE 886 AA; 99018 MW; AC0594C685257688 CRC64;

Query Match 6.7%; Score 85.5; DB 1; Length 886;
Best Local Similarity 23.8%; Pred. No. 21;
Matches 36; Conservative 34; Mismatches 62; Indels 19; Gaps 7;

QY 75 YPDHMKQHDFFKSAMPEGYVQERT-----IFFKDDGNKYKTRAEVKFEGD---TLVNRIE 125
Db 476 YFENAIL-DLTNSYPHFYDGSSETAVAGRLADSDMNFK--ADVKGHALNDLTFTEVD 532
QY 126 LKGIID--FKEDGNILGHKLEYNVYIMADKQKNGIKVNFKIRHNIEDGVSQVLADHY 183
Db 533 MKEMDAALKEQGVIFGNVIERLWAYLTIEQLLEKRN---AHGEKENLTAQALELSKY 589
QY 184 QQNTPIGDCGPVLLPDNHYLSTQSALSQDPNE 214
Db 590 HFVTPLTPMVVTKPEDN--EDQTSIADKPGE 618

RESULT 14
SY62_DISOM STANDARD; PRT; 439 AA.
ID SY62_DISOM
AC P24506;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Synaptotagmin B (Synaptic vesicle protein O-p65-B).
GN P65-B.
OS Discopysge ommata (Electric ray).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squala; Hypnosqualea; Pristiogalea; Batoidea;
OC Torpediniformes; Narcinoidei; Narcinidae; Discopysge.
OX NCBI_TaxID=7785;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=91273991; PubMed=2054189;
RA Wendland B., Miller K.G., Schilling J., Scheller R.H.;
RT "Differential expression of the p65 gene family.";
RL Neuron 6:993-1007(1991).
CC -!- FUNCTION: May have a regulatory role in the membrane interactions
CC during trafficking of synaptic vesicles at the active zone of the
CC synapse. It binds acidic phospholipids with a specificity that
CC requires the presence of both an acidic head group and a diacyl
CC backbone.
CC -!- SUBUNIT: Homodimer or homotrimer (possible).
CC -!- SUBCELLULAR LOCATION: Synaptic vesicles in neurons.
CC -!- TISSUE SPECIFICITY: Spinal cord, brainstem, midbrain and electric
CC organ.
CC -!- SIMILARITY: Contains 2 C2 domains.
CC -!- SIMILARITY: Belongs to the synaptotagmin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

EMBL; M64276; AAA49228.1; -.
PIR; JH0414; JH0414.
HSP; P21707; 1BYN.
InterPro; IPR000008; C2.
InterPro; IPR008973; C2_CaLB.
InterPro; IPR002149; LRI.
Pfam; PF00168; C2; 2.
SMART; SM00239; C2; 2.
PROSITE; PS00499; C2_DOMAIN 1; 2.
PROSITE; PS50004; C2_DOMAIN 2; 2.
Transmembrane; Repeat; Synapse; Multigene family; Glycoprotein.
DOMAIN 1 74
TRANSMEM 75 101
DOMAIN 102 439
DOMAIN 153 339
DOMAIN 173 262
DOMAIN 304 395
DOMAIN 6 6
CARBOHYD 46 46
CARBOHYD 439 AA; 49278 MW; 2033F05FDBC693F39 CRC64;
SEQUENCE

```

```

very Match      6.7%; Score 85; DB 1; Length 439;
set Local Similarity 19.2%; Pred. No. 10;
tches 55; Conservative 46; Mismatch 94; Indels 92; Gaps 12;

16 LVLELDGVNQHFSVSGEGEDATYKGLTKLICTTGKLPV-PWPLTVTGSYGV----- 69
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
38 MWPIDTGDNSTEAGVFGKND-VPEKLEKEMNELQKIPLPWPALIAIAIVSGLLLTLC 96
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
70 -----QCTSRYPDHMKQHDFFKSAPEGYVQERTIFPKDCGNYKTRAEVKEFG----- 117
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
97 CLCICKCKCKCKKKNKCKGK-----KNDINMK-----DVKGSGGNQDD 138
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
118 -----DTLVNRTELKIGIDFKEDGNI--LGHKLEYNNSH----- 149
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
139 DAETGLTEGEDKEEAEKEEKGKIQFSLDYDFQANQLTVGIIQAELPALDMGTSDDPY 198
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
150 -NVIYIMADKKQN-GIKVN-----FKIRHNIEDGSVQLA-----DHYQONTPI 189
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
199 VKVFLPLDKKKKYETVKQKTLNPTNSESFVFKVPYQELGGKTLMAVYDFDRFSKHDCI 258
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
190 GDGPFVLLPD-----NHYLSTQSALSKOPNEKRDRHMVLLGFVTAAG 229
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
190 GQVTVLMTKTVYIGQGLPEWEDIPESAKERPEKLGDICTSLRYVYPTAG 305
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

```

ULT 15  
 1\_AZOCA  
 GLNA\_AZOCA STANDARD; PRT; 468 AA.  
 P94125;  
 30-MAY-2000 (Rel. 39, Created)  
 30-MAY-2000 (Rel. 39, Last sequence update)  
 28-FEB-2003 (Rel. 41, Last annotation update)  
 Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase).  
 GLNA.  
 Azorhizobium caulinodans.  
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 Hyphomicrobiales; Azorhizobium.  
 NCBI\_TaxID=7;  
 [1]  
 SEQUENCE FROM N.A.  
 STRAIN=ORS571;  
 MEDLINE=97315229; PubMed=9171403;  
 Michel-Reydellet N., Desnoes N., Elmerich C., Kaminski P.A.;  
 "Characterization of Azorhizobium caulinodans gls and glsA genes:  
 involvement of the P(II) protein in symbiotic nitrogen fixation.";  
 J. Bacteriol. 179:3580-3587(1997).  
 -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate

L-glutamine.

-!- ENZYME REGULATION: The activity of this enzyme is controlled by adenylation. The fully adenylation enzyme complex is inactive.

-!- SUBUNIT: Oligomer of 12 subunits arranged in the form of two hexagons (By similarity).

-!- SUBCELLULAR LOCATION: Cytoplasmic.

-!- SIMILARITY: Belongs to the glutamine synthetase family.

-----

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC	EMBL; Y10213; CAA71265.1; -.
DR	HSSP; P06201; ILGR.
DR	InterPro; IPR008147; Gln_synt_beta.
DR	InterPro; IPR008146; Gln_synt_C.
DR	InterPro; IPR004809; GlnA.
DR	InterPro; IPR001637; GlnA_adenyltn_S.
DR	Pfam; PF00120; gln-synt; 1.
DR	Pfam; PF03951; gln-synt_N; 1.
DR	PDfam; PD001057; Gln_synt_C; 1.
DR	TIGRFAMs; TIGR00653; GlnA; 1.
DR	PROSITE; PS00180; GLNA_1; 1.
DR	PROSITE; PS00181; GLNA_ATP; 1.
DR	PROSITE; PS00182; GLNA_ADENYLATION; 1.
KW	Nitrogen fixation; Ligase.
FT	Binding 397
SO	SEQUENCE 468 AA; 52359 MW; 36C759B9C9D5303 CRC64; GLUTAMINE) (BY SIMILARITY).

	Query Match	6.6%;	Score 84.5;	DB 1;	Length 468;	
	Best Local Similarity	24.9%;	Pred. No.12;			
	Matches	45;	Conservative 21;	Mismatches 60;	Indels 55;	Gaps 10;
QY	37	DATYGKLTAKTICTTGKLPVPWPVTLVTGLSGVOCFSRYPDHMKOHDFFKSAMPEGIVQE	96			
		:   :   :   :		:   :   :   :		
Dd	78	DPFFSETLTVVC-----DVLEPT--TGEPYG----RDP-----RGIAKKAM--AVIQS	118			
		:   :   :   :		:   :   :   :		
QY	97	R-----TIFFKODGNKYTRAEVKFEGDILVNRIELKGIDFKEDGNI-----LGHKLEVNY	146			
		:   :   :   :		:   :   :   :		
Dd	119	TGIGDTVFPGPEAEFFIDDDVKFKADPNYTFKGLDSIELPNTGDIDYEGLNGLHRIKTG	178			
		:   :   :   :		:   :   :   :		
QY	147	NSHNV-----YTMADKQNGIKGVNFKIRHNIEDGSVQ-----LDHY	183			
		:   :   :   :		:   :   :   :		
Dd	179	GYPFPPLDSAQDMRESEMLASMAWGAKVE---KHHEVAQAHELGLKFGLVTMADHL	235			
		:   :   :   :		:   :   :   :		
QY	184	Q	184			
		:   :   :   :		:   :   :   :		
Dd	236	O	236			
		:   :   :   :		:   :   :   :		

Search completed: June 21, 2004, 15:55:19  
Job time : 8.55556 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:45:30 ; Search time 30.6667 Seconds  
(without alignments)  
2458.984 Million cell updates/sec

Title: US-09-887-784-64G  
Perfect score: 1275  
Sequence: 1 MVSKEGLFTGVVILVELD.....VLIGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL\_25.\*
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phase.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_virus.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_unclassified.\*
  - 15: sp\_rvirus.\*
  - 16: sp\_bacteriap.\*
  - 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1236	96.9	238	Q8GHE2	Q8ghe2 azotobacter
2	1233	96.7	238	Q93125	Q93125 aequorea vi
3	1231	96.5	238	Q8GHE4	Q8ghe4 azomonas ag
4	1230	96.5	238	Q8GHE3	Q8ghe3 azotobacter
5	1198	94.0	238	Q17105	Q17105 aequorea vi
6	1183	92.8	238	Q17106	Q17106 aequorea vi
7	1078	84.5	238	Q8WTC6	Q8wtc6 aequorea ma
8	1074	84.2	238	Q8WP95	Q8wp95 aequorea ma
9	1070	83.9	238	Q8WTC4	Q8wtc4 aequorea ma
10	1068	83.8	238	Q8WTC5	Q8wtc5 aequorea ma
11	1067	83.7	238	Q8WTC8	Q8wtc8 aequorea ma
12	1067	83.7	238	Q8WTC9	Q8wtc9 aequorea ma
13	1065	83.5	238	Q8WTC7	Q8wtc7 aequorea ma
14	1063	83.4	238	Q8WTC5	Q8wtc5 aequorea ma
15	252.5	19.8	225	Q95UA7	Q95ua7 montastraea
16	252.5	19.8	225	Q720W5	Q720w5 montastraea

17	247	19.4	225	5	Q963F5	Q963f5 montastraea
18	244.5	19.2	236	5	Q8TGU0	Q8tgu0 dendroneptht
19	240	18.8	225	5	Q8I6J8	Q8i6j8 trachyphyll
20	239.5	18.8	225	5	Q720W9	Q720w9 montastraea
21	238.5	18.7	266	5	Q9U6Y3	Q9u6y3 clavularia
22	233	18.3	225	5	Q720W4	Q720w4 montastraea
23	232	18.2	224	5	Q8MU48	Q8mu48 montastraea
24	229	18.0	225	5	Q8T5F1	Q8t5f1 montastraea
25	212.5	16.7	259	5	Q8MMA2	Q8mma2 agaricia fr
26	212	16.6	239	5	Q8VMA1	Q8vma1 agaricia ag
27	211	16.5	227	5	Q720W6	Q720w6 montastraea
28	211	16.5	234	5	Q720W7	Q720w7 montastraea
29	206	16.2	234	5	Q8T5F2	Q8t5f2 montastraea
30	206	16.2	234	5	Q8MU47	Q8mu47 montastraea
31	205.5	16.1	229	5	Q9U6Y6	Q9u6y6 anemonia ma
32	204.5	16.0	232	5	Q9GP15	Q9gpi5 anemonia su
33	203.5	16.0	238	5	Q9BLX9	Q9bli9 renilla mue
34	203	15.9	221	5	Q95P04	Q95p04 gonlopora t
35	203	15.9	227	5	Q962P9	Q962p9 montastraea
36	203	15.9	227	5	Q720W8	Q720w8 montastraea
37	200.5	15.7	225	5	Q9U6Y8	Q9u6y8 discosoma s
38	200.5	15.7	232	5	Q9GZ28	Q9gz28 anemonia s
39	199	15.6	227	5	Q95V70	Q95v70 montastraea
40	198.5	15.6	222	5	Q72168	Q72168 cerianthus
41	198.5	15.6	225	5	Q8TG79	Q8tg79 radianthus
42	198	15.5	235	5	Q8T5F0	Q8t5f0 scolymia cu
43	197.5	15.5	214	5	Q86LV7	Q86lv7 meandrina m
44	196.5	15.4	214	5	Q86LV8	Q86lv8 meandrina m
45	196.5	15.4	232	5	Q9U6Y7	Q9u6y7 discosoma s

ALIGNMENTS

RESULT 1

Q8GHE2	ID	Q8GHE2	PRELIMINARY;	PRT;	238 AA.
AC	Q8GHE2;				
DT	01-MAR-2003	(TrEMBLrel. 23, Created)			
DT	01-MAR-2003	(TrEMBLrel. 23, Last sequence update)			
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)			
DE	Green fluorescence protein.				
GN	2289GFP.				
OS	Azotobacter vinelandii.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;				
OC	Pseudomonadaceae; Azotobacter.				
OX	NCBI_TaxID=354;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=DSM2289;				
RA	Koranyi P., Berenyi M., Burg K.;				
RT	"Occurrence of green fluorescence protein in diazotrophic bacteria				
RT	Azomonas and Azotobacter."				
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF324408; AAN86140.1; -.				
DR	GO; GO:0006091; P:energy pathways; IEA.				
DR	InterPro; IPR0009017; GFP like.				
DR	InterPro; IPR000786; Green_fl_protein.				
DR	Pfam; PF01353; GFP; 1.				
DR	PRINTS; PR01229; GFP; 1.				
DR	ProDom; PD013756; Green_fl_protein; 1.				
SQ	SEQUENCE 238 AA; 26914 MW; F84840F1F9064018 CRC64;				

Query Match 96.9%; Score 1236; DB 2; Length 238;  
Best Local Similarity 97.5%; Pred. No. 4.9e-96;  
Matches 232; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY	2	VSKGELFTGVVILVELDGVNGHKFSVSGEGDATYKGLTKFKICTTCKLPVWPPTL 61
Db	1	MSKGELFTGVVILVELDGVNGHKFSVSGEGDATYKGLTKFKICTTCKLPVWPPTL 60
QY	62	VTGLSYGVQCFRYPDHMKQHDFFKSPMEGVYQERTIPFKDGNKYKTRAEVKEGDTLV 121

```
Db 61 VTTFSYGVQCFSRYPDHMKHDFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120
QY 122 NRIELKGIDFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFKIRHNIEDGVSQVLAD 181
Db 121 NRIELKGIDFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFKIRHNIEDGVSQVLAD 180
QY 182 HYQONTPIGDGPVLLPDNHYLSQTSALSQKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSQTSALSQKDPNEKRDHMLVGFVTAAGITLGMDELYK 238

RESULT 2
ID Q93125 PRELIMINARY; PRT; 238 AA.
AC Q93125;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein mutant 3.
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96305137; PubMed=8707053;
RA Cormack B.P., Valdivia R.H., Falkow S.;
RT "FACS-optimized mutants of the green fluorescent protein (GFP).";
RL Gene 173:33-38(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Cormack B.P., Bertam G., Egerton M., Gow N.A.R., Falkow S.,
RA Brown A.J.P.;
RT "Yeast Enhanced Green Fluorescent Protein (yEGFP): a reporter of gene
RT expression in Candida albicans.";
RL Microbiology 0:0-0(1996).
DR EMBL; U73901; AAB18957.1; -.
DR HSP; P42212; IAFP.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green fl protein; 1.
SQ SEQUENCE 238 AA; 26940 MW; A28622809ADEA60 CRC64;

Query Match 96.7%; Score 1233; DB 5; Length 238;
Best Local Similarity 97.1%; Pred. No. 8.7e-96;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKGLTKLFICTTGKLPVPMPTL 61
Db 1 MSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKGLTKLFICTTGKLPVPMPTL 60
QY 62 VTGLSYGVQCFSRYPDHMKHDFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121
Db 61 VTTFSYGVQCFSRYPDHMKHDFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120
QY 122 NRIELKGIDFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFKIRHNIEDGVSQVLAD 181
Db 121 NRIELKGIDFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFKIRHNIEDGVSQVLAD 180
QY 182 HYQONTPIGDGPVLLPDNHYLSQTSALSQKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSQTSALSQKDPNEKRDHMLVGFVTAAGITLGMDELYK 238

RESULT 3
Q8GHE4
ID Q8GHE4 PRELIMINARY; PRT; 238 AA.
AC Q8GHE4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
Query Match 96.5%; Score 1230; DB 2; Length 238;
```

```
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescence protein.
GN 375GFP.
OS Azomonas agilis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azomonas.
OX NCBI_TaxID=116849;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
RT Azomonas and Azotobacter.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324405; AAN86137.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green fl protein; 1.
SQ SEQUENCE 238 AA; 26902 MW; 15FE9B9C5B4F6B89 CRC64;

Query Match 96.5%; Score 1231; DB 2; Length 238;
Best Local Similarity 97.1%; Pred. No. 1.3e-95;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKGLTKLFICTTGKLPVPMPTL 61
Db 1 MSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKGLTKLFICTTGKLPVPMPTL 60
QY 62 VTGLSYGVQCFSRYPDHMKHDFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121
Db 61 VTTFSYGVQCFSRYPDHMKHDFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120
QY 122 NRIELKGIDFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFKIRHNIEDGVSQVLAD 181
Db 121 NRIELKGIDFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFKIRHNIEDGVSQVLAD 180
QY 182 HYQONTPIGDGPVLLPDNHYLSQTSALSQKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSQTSALSQKDPNEKRDHMLVGFVTAAGITLGMDELYK 238

RESULT 4
Q8GHE3
ID Q8GHE3 PRELIMINARY; PRT; 238 AA.
AC Q8GHE3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescence protein.
GN 85GFP.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
RT Azomonas and Azotobacter.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324406; AAN86138.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green fl protein; 1.
SQ SEQUENCE 238 AA; 26887 MW; E0E1616BD2AF6188 CRC64;

Query Match 96.5%; Score 1230; DB 2; Length 238;
```



```
Best Local Similarity 97.1%; Pred. No. 1.6e-95;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTLKFICTTGKLPVWPPTL 60

QY 62 VTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKRHDFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKVNFKIRNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKVNFKIRNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMYLLGFVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMYLLGFVTAAGITLGMDELYK 238

RESULT 5
Q17105
ID Q17105 PRELIMINARY; PRT; 238 AA.
AC Q17105;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83960; CAA58789.1; -.
DR PIR; JS0692; JQ1514.
DR HSSP; P42212; 1BFP.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748E44 CRC64;

Query Match 94.0%; Score 1198; DB 5; Length 238;
Best Local Similarity 93.7%; Pred. No. 7.7e-93;
Matches 223; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGDVNGKFSQSGEGDATYKGLTLNFICTTGKLPVWPPTL 60

QY 62 VTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKVNFKIRNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHKMEYNNSHVYIMGDKPKNGIKVNFKIRNIDKGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMYLLGFVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMYLLGFVTAARITHGMDELYK 238

RESULT 6
Q17106
ID Q17106 PRELIMINARY; PRT; 238 AA.
AC Q17106;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83959; CAA58789.1; -.
DR PIR; JS0692; JQ1514.
DR HSSP; P42212; 1GFL.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748E44 CRC64;

Query Match 92.8%; Score 1183; DB 5; Length 238;
Best Local Similarity 92.9%; Pred. No. 1.4e-91;
Matches 221; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGDVNGKFSVRGEGEGDATYKGLTLKFICTTGKLPVWPPTL 60

QY 62 VTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKVNFKIRNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHKMEYNNSHVYIMGDKPKNGIKVNFKIRNIDKGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMYLLGFVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPHGRDHMYLLGFVTSAGITHGMDELYK 238

RESULT 7
Q8WTC6
ID Q8WTC6 PRELIMINARY; PRT; 238 AA.
AC Q8WTC6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN-GFPxm19uv;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RA "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435431; AAL33916.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP_like.
DR Pfam; PF01353; GFP; 1.
```

```
DR PRINTS; PRO1229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27015 MW; 6B8FD75E88926903 CRC64;

Query Match      84.5%; Score 1078; DB 5; Length 238;
Best Local Similarity 81.9%; Pred. No. 9.6e-83;
Matches 195; Conservative 20; Mismatches 23; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTKLFTCTTGKLPVWPPTL 61
   :|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 1 MSKGEELFTGIVPVLIELDGDVHGHKFSVRGEGDADYKLEIKFICTTGKLPVWPPTL 60

QY 62 VTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
   :|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 61 VTTLSYGIQCFARYPEHMKWMDFFKSAMPEGYIQTERTIFFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRIELKGIDFEDGNILGHKLEYNNSHVYIMADKQNGIKVNFIRHNIEDGSVOLAD 181
   :|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 121 NRIELKGMDFKEDGNILGHKLEYNNSHVYIMPDKANGLKVNFKIRHNIEGGVQLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALSQSDPNKRDHMLVLFVTAAGITLGMDELYK 239
   :|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 181 HYQTNVPLGDGPVLIPINHVLSTQTALSKDRNETRDHMLVLEFFSACGHTGMDLYK 238

RESULT 8
Q8WP95 PRELIMINARY; PRT; 238 AA.
AC Q8WP95;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DR GO; GO:0006091; P:energy pathways; IEA.
DE Green fluorescent protein.
GN GFPXM.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPxm161;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qian Y.X., Pang S.Q.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY013824; AAK02059.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PRO1229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27049 MW; 8185D05E529012B CRC64;

Query Match      84.2%; Score 1074; DB 5; Length 238;
Best Local Similarity 81.5%; Pred. No. 2.1e-82;
Matches 194; Conservative 20; Mismatches 24; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTKLFTCTTGKLPVWPPTL 61
   :|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 1 MSKGEELFTGIVPVLIELDGDVHGHKFSVRGEGDADYKLEIKFICTTGKLPVWPPTL 60

QY 62 VTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
   :|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 61 VTTFSYGIQCFARYPEHMKWMDFFKSAMPEGYIQTERTIFFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRIELKGIDFEDGNILGHKLEYNNSHVYIMADKQNGIKVNFIRHNIEDGSVOLAD 181
   :|||||:|:|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|
Db 121 NRIELKGMDFKEDGNILGHKLEYNNSHVYIMPDKANGLKVNFKIRHNIEGGVQLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALSQSDPNKRDHMLVLFVTAAGITLGMDELYK 239
   :|||||:|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|
Db 181 HYQTNVPLGDGPVLIPINHVLSTQTALSKDRNETRDHMLVLEFFSACGHTGMDLYK 238
```

```
RESULT 9
Q8WTC4 PRELIMINARY; PRT; 238 AA.
AC Q8WTC4;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Shg24;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL33918.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PRO1229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match      83.9%; Score 1070; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 4.5e-82;
Matches 196; Conservative 17; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKGLTKLFTCTTGKLPVWPPTL 61
   :|||||:|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|
Db 1 MSKGEELFTGVVPILVELDGDVHGHKFSVRGEGDADYKLEIKFICTTGKLPVWPPTL 60

QY 62 VTGLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
   :|||||:|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|
Db 61 VTTLGILGILCFARYPEHMKQNDFFKSAMPEGYIQTERTIFFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRIELKGIDFEDGNILGHKLEYNNSHVYIMADKQNGIKVNFIRHNIEDGSVOLAD 181
   :|||||:|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|
Db 121 NRIELKGMDFKEDGNILGHKLEYNNSHVYIMPDKANGLKVNFKIRHNIEGGVQLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALSQSDPNKRDHMLVLFVTAAGITLGMDELYK 239
   :|||||:|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|
Db 181 HYQTNVPLGDGPVLIPINHVLSTQTALSKDRNETRDHMLVLEFFSACGHTGMDLYK 238

RESULT 10
Q8WTD0 PRELIMINARY; PRT; 238 AA.
AC Q8WTD0;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPxm161;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla.";
```

```
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435427; AAL33912.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26997 MW; 5F80A192173CB84D CRC64;

Query Match      83.7%; Score 1067; DB 5; Length 238;
Best Local Similarity 81.1%; Pred. No. 8.1e-82;
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKLFTCTTGKLPVWPPTL 61
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSKGEELFTGIVPVLIELDGDVGHKFSVRGEGDADYKGLKIKFTCTTGKLPVWPPTL 60

QY 62 VTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTLGYGIOCFARYPEHMKMNDFFKSAPEGYVQERTIFFQDDGKYKTRGEVKEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNHVIYIMADKQKNGIKVNFKIRHNIEDGSVOLAD 181
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 NRIELKGMDPKEDGNILGHKLEYNFNHSHVYIMPDKANNGKLVNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 HYQTNVPLGDGPVLLIPINHYLSFQTALSKDRNETRDHMFVLEFFSACGHTGMDLYK 238

RESULT 12
Q8WTC9 PRELIMINARY; PRT; 238 AA.
ID Q8WTC9
AC Q8WTC9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPxm163;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435428; AAL33913.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27031 MW; 5F80A19C19DC584D CRC64;

Query Match      83.7%; Score 1067; DB 5; Length 238;
Best Local Similarity 81.1%; Pred. No. 8.1e-82;
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKLFTCTTGKLPVWPPTL 61
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGIVPVLIELDGDVGHKFSVRGEGDADYKGLKIKFTCTTGKLPVWPPTL 60

QY 62 VTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTLGYGIOCFARYPEHMKMNDFFKSAPEGYVQERTIFFQDDGKYKTRGEVKEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNHVIYIMADKQKNGIKVNFKIRHNIEDGSVOLAD 181
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 NRIELKGMDPKEDGNILGHKLEYNFNHSHVYIMPDKANNGKLVNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 HYQTNVPLGDGPVLLIPINHYLSFQTALSKDRNETRDHMFVLEFFSACGHTGMDLYK 238

RESULT 13
Q8WTC7 PRELIMINARY; PRT; 238 AA.
ID Q8WTC7
AC Q8WTC7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435427; AAL33912.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26997 MW; 5F80A192173CB84D CRC64;

Query Match      83.7%; Score 1067; DB 5; Length 238;
Best Local Similarity 81.1%; Pred. No. 8.1e-82;
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKLFTCTTGKLPVWPPTL 61
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGIVPVLIELDGDVGHKFSVRGEGDADYKGLKIKFTCTTGKLPVWPPTL 60

QY 62 VTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTLGYGIOCFARYPEHMKMNDFFKSAPEGYVQERTIFFQDDGKYKTRGEVKEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNHVIYIMADKQKNGIKVNFKIRHNIEDGSVOLAD 181
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 NRIELKGMDPKEDGNILGHKLEYNFNHSHVYIMPDKANNGKLVNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 HYQTNVPLGDGPVLLIPINHYLSFQTALSKDRNETRDHMFVLEFFSACGHTGMDLYK 238

RESULT 14
Q8WTC8 PRELIMINARY; PRT; 238 AA.
ID Q8WTC8
AC Q8WTC8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPxm163;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435429; AAL33914.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;

Query Match      83.7%; Score 1067; DB 5; Length 238;
Best Local Similarity 81.1%; Pred. No. 8.1e-82;
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKLFTCTTGKLPVWPPTL 61
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGIVPVLIELDGDVGHKFSVRGEGDADYKGLKIKFTCTTGKLPVWPPTL 60

QY 62 VTGLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTLGYGIOCFARYPEHMKMNDFFKSAPEGYVQERTIFFQDDGKYKTRGEVKEGDTLV 120
```

```

RN 1 MSKGEELFTGVVPILVELDGDVHGKFSVRGEGGDADYKLEIKFICTTGKLPVWPPTL 60
RP SEQUENCE FROM N.A.
RC STRAIN=GFPM191uv;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435430; AAL33915.1; -;
DR GO: GO:0006091; P:energy pathways; IEA.
DR InterPro: IPR009017; GFP_like.
DR DR InterPro: IPR000786; Green_fl_protein.
DR DR Pfam: PF01353; GFP; 1.
DR PRINTS: PR01229; GFP.
DR PRODOM: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27002 MW; 8D5BA2982264C018 CRC64;

Query Match 83.5%; Score 1065; DB 5; Length 238;
Best Local Similarity 81.1%; Pred. No. 1.2e-81;
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVHGKFSVRGEGGDADYKLEIKFICTTGKLPVWPPTL 61
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MSKGEELFTGVVPILVELDGDVHGKFSVRGEGGDADYKLEIKFICTTGKLPVWPPTL 60
QY 52 VTGLSYGVOCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNKYNKTRAEVKFEGDTLV 121
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 VTTLSYGILCFARYPEHMKQNDFFKSAPEGYIQERTIFFQDDGKYKTGEVKFEGDTLV 120
QY 122 NRIELKGIDPKEDGNILGHKLEYNNSHNHYIMADKQNGIKVNFIRHNIEDGSVQLAD 181
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 NRIELKGIDPKEDGNILGHKLEYNNSHNHYIMADKQNGIKVNFIRHNIEDGSVQLAD 180
QY 182 HYQONTPIGDGPVLLPDNHYLSQTSALSADPNKRDHMLLGFVTAAGITLGMDELYK 239
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 HYQONTPIGDGPVLLPDNHYLSQTSALSADPNKRDHMLLGFVTAAGITLGMDELYK 238

RESULT 14
Q8WTC5 PRELIMINARY; PRT; 238 AA.
ID Q8WTC5
AC Q8WTC5
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Orange fluorescent protein.
GN Aequorea macrodactyla.
OS Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OPPM;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435432; AAL33917.1; -;
DR GO: GO:0006091; P:energy pathways; IEA.
DR InterPro: IPR009017; GFP like.
DR DR InterPro: IPR000786; Green_fl_protein.
DR DR Pfam: PF01353; GFP; 1.
DR PRINTS: PR01229; GFP.
DR PRODOM: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27018 MW; 75521EA5534E573A CRC64;

Query Match 83.4%; Score 1063; DB 5; Length 238;
Best Local Similarity 81.9%; Pred. No. 1.8e-81;
Matches 195; Conservative 17; Mismatches 26; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVHGKFSVRGEGGDADYKLEIKFICTTGKLPVWPPTL 61

```

```

Db 1 MSKGEELFTGVVPILVELDGDVHGKFSVRGEGGDADYKLEIKFICTTGKLPVWPPTL 60
QY 62 VTGLSYGVOCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNKYNKTRAEVKFEGDTLV 121
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 VTTLSYGILCFARYPEHMKQNDFFKSAPEGYIQERTIFFQDDGKYKTGEVKFEGDTLV 120
QY 122 NRIELKGIDPKEDGNILGHKLEYNNSHNHYIMADKQNGIKVNFIRHNIEDGSVQLAD 181
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 NRIELKGIDPKEDGNILGHKLEYNNSHNHYIMADKQNGIKVNFIRHNIEDGSVQLAD 180
QY 182 HYQONTPIGDGPVLLPDNHYLSQTSALSADPNKRDHMLLGFVTAAGITLGMDELYK 239
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 HYQONTPIGDGPVLLPDNHYLSQTSALSADPNKRDHMLLGFVTAAGITLGMDELYK 238

RESULT 15
Q95UA7 PRELIMINARY; PRT; 225 AA.
ID Q95UA7
AC Q95UA7
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Cyan fluorescent protein (Fragment).
OS Montastraea cavernosa (great star coral).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastraea.
OX NCBI_TaxID=63558;
RN [1]
RP SEQUENCE FROM N.A.
RA Falkowski P.G., Sun Y.;
RT "Montastraea cavernosa fluorescent protein.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY056460; AAL17905.1; -;
DR GO: GO:0006091; P:energy pathways; IEA.
DR InterPro: IPR009017; GFP like.
DR DR InterPro: IPR000786; Green_fl_protein.
DR DR Pfam: PF01353; GFP; 1.
DR PRINTS: PR01229; GFP.
DR PRODOM: PD013756; Green_fl_protein; 1.
FT NON_TER
SQ SEQUENCE 225 AA; 25775 MW; 52DE2F716D083524 CRC64;

Query Match 19.8%; Score 252.5; DB 5; Length 225;
Best Local Similarity 31.1%; Pred. No. 2.6e-13;
Matches 64; Conservative 43; Mismatches 82; Indels 17; Gaps 7;

QY 12 VPIVELDGDVNGHKFSVRGEGGDADYKLEIKFICTTGKLPVWPPTLVGLSYGVQ 70
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7 VMKIKLRMDGIVNGHKFMITGEGEGKPFEGTHTIILKVXEGGLPPAYDILTATFQYGNR 66
QY 71 CFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNKYNKTRAEVKFEGDTLVNRIELKGID 130
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 VFTKYPRDIP--DYFKQSPFEGYSWERSMTFEDQGYCTVTSIDIKLEGDCFFYIRFVGN 124
QY 131 FKEDGNILGHKLEYNNSHNHYIMADKQNGIKVNFIRHNIEDGSVQLADHYQONTPI 189
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
125 FPGSGPVMQKTKLWEPSTENMVV-----RDGVLGDVSRITLLEGD----KHRCNFRS 175
QY 190 GDGP-----VLLPDNHYLSQTSALSADPNKRDHMLLGFVTAAGITLGMDELYK 211
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
176 TYGAKKGVVLPYHYFVDHRIEILSHD 201

Search completed: June 21, 2004, 16:00:08
Job time : 32.7778 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:41:49 ; Search time 47.1111 Seconds  
(without alignments)  
1433.395 Million cell updates/sec

Title: US-09-887-784-64L

Perfect score: 1273  
Sequence: 1 MYSKGELFTGVVILVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:.\*  
1: geneseqp1980s:.\*  
2: geneseqp1990s:.\*  
3: geneseqp2000s:.\*  
4: geneseqp2001s:.\*  
5: geneseqp2002s:.\*  
6: geneseqp2003as:.\*  
7: geneseqp2003bs:.\*  
8: geneseqp2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1268	99.6	239	5 AAE17518	Aae17518 Enhanced
2	1268	99.6	363	6 ABR40352	Abr40352 Human ami
3	1268	99.6	893	4 AAG65781	Aag65781 Amino aci
4	1268	99.6	1132	4 AAG55782	Aag55782 Amino aci
5	1260	99.0	239	5 AAE17517	Aae17517 Enhanced
6	1257	98.7	239	3 AAB22882	Aab22882 Enhanced
7	1257	98.7	239	3 AAY54349	Aay54349 Amino aci
8	1257	98.7	239	3 AAY79584	Aay79584 EGFP sign
9	1257	98.7	239	4 AAB50804	Aab50804 Jellyfish
10	1257	98.7	239	4 AAB55900	Aab55900 A. victor
11	1257	98.7	239	4 AAB31171	Aab31171 Amino aci
12	1257	98.7	239	5 AAG66198	Aag66198 A. victor
13	1257	98.7	239	5 AAG94444	Aag94444 Protease
14	1257	98.7	239	5 AAE14599	Aae14599 Aequorea
15	1257	98.7	239	6 AAE34958	Aae34958 Aequorea
16	1257	98.7	239	6 AAG79829	Aag79829 Green flu
17	1257	98.7	239	6 ABR83616	Abr83616 Green flu
18	1257	98.7	239	6 ADA38074	Ada38074 Aequorea
19	1257	98.7	239	7 ABUG3204	Abug3204 Aequorea
20	1257	98.7	239	7 ADC18358	Adc18358 EGFP (enh
21	1257	98.7	239	7 ABW00914	Abw00914 Aequorea
22	1257	98.7	239	7 ADE28570	Ade28570 Enhanced
23	1257	98.7	246	7 ABM79011	Abm79011 Enhanced
24	1257	98.7	248	5 AAG68319	Aag68319 Jellyfish
25	1257	98.7	259	5 AAU99804	Aau99804 Biomembra

26	1257	98.7	265	2 AAW97451	Aaw97451 Wild-type
27	1257	98.7	268	5 AAU99803	Aau99803 Biomembra
28	1257	98.7	270	5 AAU99802	Aau99802 Biomembra
29	1257	98.7	272	5 AAU99800	Aau99800 Biomembra
30	1257	98.7	273	5 AAU99801	Aau99801 Biomembra
31	1257	98.7	280	5 AAU99807	Aau99807 Biomembra
32	1257	98.7	281	3 AAY50142	Aay50142 Green flu
33	1257	98.7	281	3 AAE24252	Aae24252 EGFP-MODC
34	1257	98.7	281	5 AAU10888	Aau10888 EGFP-MODC
35	1257	98.7	286	7 ADE28562	Ade28562 EGFP/ hum
36	1257	98.7	289	7 ADE28564	Ade28564 EGFP/ hum
37	1257	98.7	290	7 ADE28568	Ade28568 EGFP/ hum
38	1257	98.7	290	7 ADE28566	Ade28566 EGFP/ hum
39	1257	98.7	294	3 AAB22860	Aab22860 GFP-DEVD-
40	1257	98.7	294	3 AAY79638	Aay79638 Caspase-3
41	1257	98.7	294	5 ABG94422	Abg94422 Recombina
42	1257	98.7	308	2 AAY42181	Aay42181 EGFP/DRM
43	1257	98.7	320	6 ABR83620	Abr83620 HUB1-GFP
44	1257	98.7	323	3 AAY54359	Aay54359 GFP mutan
45	1257	98.7	323	6 ABR83621	Abr83621 HUB1-GFP

ALIGNMENTS

RESULT 1  
AAE17518  
ID AAE17518 standard; protein; 239 AA.  
AC AAE17518;  
XX  
XX  
DT 22-APR-2002 (first entry)  
XX  
DE Enhanced F64L-E222G jellyfish green fluorescent protein mutant.  
XX  
KW Jellyfish; green fluorescent protein; GFP; protein redistribution;  
KW cellular function; genetic reporter; mutant; Stoke's shift; mutein.  
XX  
OS Aequorea victoria.  
OS Synthetic.  
XX  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 65 /note= "Wild type Phe substituted with Leu; This  
FT corresponds to position 64 in the wild type protein"  
FT  
FT Misc-difference 223 /note= "Wild type Glu substituted with Gly; This  
FT corresponds to position 222 in the wild type protein"  
FT  
XX WO200198338-A2.  
XX  
PD 27-DEC-2001.  
XX  
XX 18-JUN-2001; 2001WO-BF006848.  
XX  
XX 19-JUN-2000; 2000DK-00000953.  
XX  
XX 20-JUN-2000; 2000US-0212681P.  
XX  
XX 10-MAY-2001; 2001DK-00000739.  
XX  
XX 10-MAY-2001; 2001US-0290170P.  
XX  
XX (BIOI-) BIOIMAGE AS.  
XX  
XX Bjorn SP, Pagliaro L, Thastrup O;  
XX  
XX WPI; 2002-098224/13.  
XX  
XX N-PSDB; AAB28163.  
XX  
XX Novel fluorescent protein in in vitro assay for measuring protein kinase  
XX activity or dephosphorylation activity, or for measuring protein  
XX redistribution, has a green fluorescent protein with F64L and E222G  
XX mutation.  
XX  
XX Claim 9; Page 37; 41pp; English.  
PS

XX The invention relates to a fluorescent protein derived from green  
 CC fluorescent protein (GFP) or its analogue. The GFP containing mutations  
 CC at F64L and E222G has a bigger compared to other GFP's making it very  
 CC suitable for high throughput screening due to better resolution. The  
 CC fluorescent protein is useful in *in vitro* assays for measuring protein  
 CC kinase activity or dephosphorylation activity, or for measuring protein  
 CC redistribution. The fluorescent protein is useful in studying cellular  
 CC functions in living cells; as protein tags in transgenic animals, living  
 CC and fixed cells; organelle tags, secretion marker and genetic reporter.  
 CC The fluorescent protein is also useful as a cell or organelle integrity  
 CC marker, a marker for changes in cell morphology, as transfection marker,  
 CC and as a marker to be used in combination with fluorescence activated  
 CC cell sorting (FACS). The novel proteins can also be used as reporters to  
 CC monitor live or dead biomass of organisms, such as fungi. The fluorescent  
 CC protein is also useful as markers in transcriptional and translational  
 CC fusions for performing transposon vector mutagenesis and as a reporter  
 CC for bacterial detection. Transposons encoding the fluorescent protein are  
 CC useful for screening promoters and for tagging plasmids and chromosomes.  
 CC The fluorescent protein engineered into the genome of a phage is useful  
 CC for designing diagnostic tool. The present sequence is a DNA encoding  
 CC enhanced F64L-E222G jellyfish green fluorescent protein (GFP) mutant  
 XX  
 SQ Sequence 239 AA;

Query Match 99.6%; Score 1268; DB 5; Length 239;  
 Best Local Similarity 99.8%; Pred. No. 2e-122; Indels 0; Gaps 0;  
 Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYKGLTKLFICTTGKLPVPWPT 60  
 DB 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYKGLTKLFICTTGKLPVPWPT 60  
 QY 51 LVTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 51 LVTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 QY 131 DHYQNTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
 DB 131 DHYQNTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 2  
 ABR40352  
 ID ABR40352 standard; protein; 363 AA.  
 AC ABR40352;  
 XX  
 DT 08-JUL-2003 (first entry)  
 DE Human amino acid sequence SEQ ID NO: 6.  
 DE Human; heterologous conjugate; intracellular protein.  
 KW  
 XX Homo sapiens.  
 OS Aequoria victoria.  
 OS WO2003029827-A2.  
 PN  
 XX 10-APR-2003.  
 PD  
 XX 01-OCT-2002; 2002WO-DK000651.  
 PF  
 XX 01-OCT-2001; 2001DK-00001433.  
 PR 11-OCT-2001; 2001US-0328896P.  
 XX  
 XX (BIOL-) BIOMAGE AS.  
 PA  
 XX Terry BR, Nielsen SJ;  
 PI

XX WPI: 2003-430211/40.  
 DR N-PSDB; ACC72604.  
 XX  
 PT Novel cell for identifying modulators of protein interaction, contains a  
 PT first conjugate comprising anchor protein, second conjugate having type B  
 PT interactor protein and third conjugate with detectable group.  
 XX  
 PS Disclosure; Page 112-113; 118pp; English.  
 XX  
 CC The invention relates to a novel cell, comprising three heterologous  
 CC conjugates (HC), a first HC (HC1) comprising an anchor protein that  
 CC specifically binds to an internal structure within the cell conjugated to  
 CC an interactor protein (IP) of type A, a second HC (HC2) comprising IP of  
 CC type B conjugated to a first protein of interest, and a third HC (HC3)  
 CC comprising a second protein of interest conjugated to detectable group.  
 CC The cell is useful for detecting if a compound disrupts or induces the  
 CC interaction between two intracellular proteins. The cell is also useful  
 CC for screening compounds that modulate the interaction between two  
 CC intracellular proteins. The present sequence is used in the  
 CC exemplification of the invention  
 XX  
 SQ Sequence 363 AA;

Query Match 99.6%; Score 1268; DB 6; Length 363;  
 Best Local Similarity 99.6%; Pred. No. 3.7e-122;  
 Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYKGLTKLFICTTGKLPVPWPT 60  
 DB 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYKGLTKLFICTTGKLPVPWPT 60  
 QY 61 LVTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 QY 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
 DB 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 3  
 AAG65781  
 ID AAG65781 standard; protein; 893 AA.  
 XX  
 AC AAG65781;  
 XX  
 DT 07-JAN-2002 (first entry)  
 DE Amino acid sequence of HSPDE4A1-E222G fusion protein.  
 DE PDE4; central nervous system; antinflammatory; cytostatic; nootropic;  
 KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;  
 KW fusion protein.  
 XX  
 OS Homo sapiens.  
 OS Aequorea victoria.  
 XX  
 PN WO200179526-A2.  
 XX  
 PD 25-OCT-2001.  
 XX  
 PF 11-APR-2001; 2001WO-DK000264.  
 XX  
 PR 17-APR-2000; 2000DK-00000651.  
 PR 29-MAY-2000; 2000DK-00000849.  
 XX  
 XX (BIOI-) BIOIMAGE AS.  
 PA  
 XX

PI	Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;	PI	25-OCT-2001.
PI	Praestegaard M;	XX	
XX		XX	
DR	WPI: 2001-611727/70.	XX	11-APR-2001; 2001WO-DK000264.
DR	N-PSDB; AA166852.	XX	
XX		PR	17-APR-2000; 2000DK-00000651.
XX		PR	29-MAY-2000; 2000DK-00000849.
PT	Determining if a compound is a dislocator of PDE4 for identifying compounds for treating CNS and inflammatory disease comprises identifying compounds which remove PDE4 spots.	XX	
PT		PA	(BIOI-) BIOIMAGE AS.
XX		XX	
XX		PI	Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
PS	Example 1; Page 156-160; 160pp; English.	PI	Praestegaard M;
XX		XX	
CC	The invention relates to determining, if a compound, is a dislocator of PDE4. The method comprises testing if the compound removes PDE4 spots, which may optionally be induced by a Rolipram-like reference compound, and testing if it inhibits the catalytic activity of the PDE4, where the compound is a dislocator of PDE4, if it removes PDE spots and if it does not inhibit the catalytic activity of PDE4. The method is useful for identifying compounds useful for the treatment of diseases of the central nervous system such as depression and for the treatment of inflammatory disease such as joint inflammation, Crohn's disease, inflammatory bowel disease, respiratory diseases, chronic obstructive pulmonary disease (COPD), including asthma, chronic bronchitis, pulmonary emphysema, endotoxemic shock, toxic shock syndrome, systemic lupus erythematosus, psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV infection. The use of a reagent that can mimic or reverse the effect of the compound with affinity for the catalytic site on intracellular distribution of the PDE for the preparation of a medicament. The present sequence represents the amino acid sequence of a HSPDE4A1-E222G fusion protein	XX	
XX		XX	
SQ	Sequence 893 AA;	XX	
	Query Match 99.6%; Score 1268; DB 4; Length 893;		
	Best Local Similarity 99.6%; Pred. No. 1.4e-121;		
	Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
QY	1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60	QY	1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
DB	655 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 714	DB	894 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 953
QY	61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120	QY	61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB	715 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 774	DB	954 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 1013
QY	121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180	QY	121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB	775 VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 834	DB	1014 VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 1073
QY	181 DHYQONTPIGDGPFVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239	QY	181 DHYQONTPIGDGPFVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
DB	835 DHYQONTPIGDGPFVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 893	DB	1074 DHYQONTPIGDGPFVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 1132
RESULT 4		RESULT 5	
AAAG65782		AAAE17517	
ID	AAG65782 standard; protein; 1132 AA.	ID	AAAE17517 standard; protein; 239 AA.
XX		XX	
AC	AAG65782;	AC	AAAE17517;
XX		XX	
DT	07-JAN-2002 (first entry)	DT	22-APR-2002 (first entry)
XX		XX	
DE	Amino acid sequence of HSPDE4A4-E222G fusion protein.	XX	Enhanced F64L jellyfish green fluorescent protein mutant.
XX		XX	
KW	PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;		
KW	autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;		
KW	fusion protein.		
OS	Homo sapiens.		
OS	Aequorea victoria.		
XX			
PN	WO200179526-A2.		
XX			

KW Jellyfish; green fluorescent protein; GFP; protein redistribution;  
KW cellular function; genetic reporter; mutant; Stoke's shift; mutin.  
XX  
OS Aequorea victoria.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 65  
FT /note= "Wild type Phe substituted with Leu; This  
FT corresponds to position 64 in the wild type protein"  
XX  
XX  
PN WO200198338-A2.  
XX  
XX 27-DEC-2001.  
XX  
XX 18-JUN-2001; 2001WO-EP006848.  
XX  
XX 19-JUN-2000; 2000DK-00000953.  
PR 20-JUN-2000; 2000US-0212681P.  
PR 10-MAY-2001; 2001DK-00000739.  
PR 10-MAY-2001; 2001US-0290170P.  
XX  
XX (BIOI-) BIOIMAGE AS.  
PA  
XX Bjorn .3P, Pagliaro L, Thastrup O;  
XX  
XX WPI; 2002-098224/13.  
DR N-PSDB; AAD28162.  
XX  
XX Novel fluorescent protein in in vitro assay for measuring protein kinase  
PT activity or dephosphorylation activity, or for measuring protein  
PT redistribution, has a green fluorescent protein with F64L and E222G  
PT mutation.  
XX  
XX Example 1; Page 35; 41pp; English.  
XX  
XX The invention relates to a fluorescent protein derived from green  
CC fluorescent protein (GFP) or its analogue. The GFP containing mutations  
CC at F64L and E222G has a bigger compared to other GFP's making it very  
CC suitable for high throughput screening due to better resolution. The  
CC fluorescent protein is useful in invitro assays for measuring protein  
CC kinase activity or dephosphorylation activity, or for measuring protein  
CC redistribution. The fluorescent protein is useful in studying cellular  
CC functions in living cells; as protein tags in transgenic animals, living  
CC and fixed cells; organelle tags, secretion marker and genetic reporter.  
CC The fluorescent protein is also useful as a cell or organelle integrity  
CC marker, a marker for changes in cell morphology, as transfection marker,  
CC and as a marker to be used in combination with fluorescence activated  
CC cell sorting (FACS). The novel proteins can also be used as reporters to  
CC monitor live or dead biomass of organisms, such as fungi. The fluorescent  
CC protein is also useful as markers in transcriptional and translational  
CC fusions for performing transposon vector mutagenesis and as a reporter  
CC for bacterial detection. Transposons encoding the fluorescent protein are  
CC useful for screening promoters and for tagging plasmids and chromosomes.  
CC The fluorescent protein engineered into the genome of a phage is useful  
CC for designing diagnostic tool. The present sequence is enhanced F64L  
CC jellyfish green fluorescent protein (GFP) mutant  
XX  
SQ Sequence 239 AA;  
  
Query Match 99.0%; Score 1260; DB 5; Length 239;  
Best Local Similarity 99.2%; Pred. No. 1.3e-121;  
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 MVSGBELFTGVVPIVLVDGDNVNGHKFSVSGEGDGYGKLTLPKICTTGKLPVPWPT 60  
|||||  
DB 1 MVSGBELFTGVVPIVLVDGDNVNGHKFSVSGEGDGYGKLTLPKICTTGKLPVPWPT 60  
|||||  
QY 51 LVTLISYGVQCFGRYPDPMKQHDFFKSAEMPEGVGVQERTIFFKDDGNKTRAEVKFSGDTL 120  
|||||  
DB 51 LVTLISYGVQCFGRYPDPMKQHDFFKSAEMPEGVGVQERTIFFKDDGNKTRAEVKFSGDTL 120  
|||||  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNVNSHNVIADKQNGIKVNFKIRHNIEDSGVQLA 180  
|||||



CC biosensors of the invention can be used to investigate a wide range of  
CC cellular activities and to screen compounds which modulate these  
CC activities. Biosensors containing a recognition site for caspase, for  
CC example, may be used for the screening of compounds which modulate  
CC apoptosis, while biosensors containing other protease recognition sites  
CC may be used for the detection of proteolytic toxins (such as anthrax  
CC lethal factor). The method provides improved target validation and  
CC candidate compound optimisation by combining many cell screening formats  
CC with fluorescence-based molecular reagents and computer-based feature  
CC extraction, data analysis and automation, resulting in increased quantity  
CC and speed of data collection and faster evaluation of drug candidates.  
CC Sequences AAB22881-B22885 represent fluorescent proteins which may be used  
CC as components of biosensor fusion proteins of the invention  
XX  
SQ Sequence 239 AA;

Query Match 98.7%; Score 1257; DB 3; Length 239;  
Best Local Similarity 98.7%; Pred. No. 2.7e-121;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MYSKGEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGLTKLFICTTGKLPVPWPT 60  
DB 1 MYSKGEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGLTKLFICTTGKLPVPWPT 60  
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRILKGIDFKEDGNILGHKLEYNHNVMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRILKGIDFKEDGNILGHKLEYNHNVMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQQTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLFVTAAGITLGMDELYK 239  
DB 181 DHYQQTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLFVTAAGITLGMDELYK 239

RESULT 7  
AA54349  
ID AAY54349 standard; protein; 239 AA.  
XX AAY54349;  
XX  
XX  
XX 06-APR-2000 (first entry)  
XX  
DE Amino acid sequence of the mutant green fluorescent protein EGFP.  
XX  
XX Fluorescent protein; green fluorescent protein; emission intensity;  
KW fluorescence; pH detection; pH sensor; EGFP.  
XX  
XX Synthetic.  
OS Aequorea victoria.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 65  
FT /note= "wild type Phe substituted with Leu"  
FT Misc-difference 66  
FT /note= "wild type Ser substituted with Thr"  
FT Misc-difference 232  
FT /note= "wild type His substituted with Leu"  
XX  
XX WO9964592-A2.  
XX  
PD 16-DEC-1999.  
XX  
XX 08-JUN-1999; 99WO-US012850.  
XX  
XX 09-JUN-1998; 98US-00094359.  
PR 13-OCT-1998; 98US-00172063.  
XX  
XX (REG ) UNIV CALIFORNIA.  
PA (UYOR-) UNIV OREGON STATE.  
XX

PI Tsien RY, Llopis J, Wachtter RM;  
XX  
DR WPI; 2000-116540/10.  
DR N-PSDB; AA245642.  
XX  
XX New functional engineered green fluorescent proteins, used for measuring  
FT the pH in biological samples and cells.  
XX  
PS Disclosure; Page 9; 89pp; English.  
XX  
CC The present sequence represents a functional engineered fluorescent  
CC protein based on the Aequorea green fluorescent protein (GFP). The  
CC emission intensity changes as pH varies between 5 and 10 of the present  
CC protein are novel. The functional engineered fluorescent proteins show  
CC reversible changes in fluorescence over physiological pH ranges. They can  
CC be used for determining the pH of samples and cells. The polynucleotides  
CC can also be used to produce transgenic animals. The fluorescent protein  
CC pH sensors can be delivered to cells in the form of polynucleotides  
CC encoding the protein sensor fused to a targeting signal. The targeting  
CC signal directs the expression of the protein sensors to restricted cell  
CC locations. This makes it possible to measure the pH of a precisely  
CC defined cellular region or organelle  
XX  
SQ Sequence 239 AA;

Query Match 98.7%; Score 1257; DB 3; Length 239;  
Best Local Similarity 98.7%; Pred. No. 2.7e-121;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MYSKGEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGLTKLFICTTGKLPVPWPT 60  
DB 1 MYSKGEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGLTKLFICTTGKLPVPWPT 60  
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRILKGIDFKEDGNILGHKLEYNHNVMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRILKGIDFKEDGNILGHKLEYNHNVMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQQTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLFVTAAGITLGMDELYK 239  
DB 181 DHYQQTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLFVTAAGITLGMDELYK 239

RESULT 8  
AA54349  
ID AAY79584 standard; peptide; 239 AA.  
XX  
XX AAY79584;  
XX  
XX 29-AUG-2000 (first entry)  
XX  
DE EGFP signal domain.  
XX  
XX Protease; biosensor; EGFP; signal peptide; cell screening; assay;  
KW analysis; drug discovery.  
XX  
OS Unidentified.  
XX  
XX WO200026408-A2.  
XX  
XX 11-MAY-2000.  
XX  
XX 29-OCT-1999; 99WO-US025431.  
PF  
XX  
XX 30-OCT-1998; 98US-0106308P.  
PR 26-MAY-1999; 99US-0136078P.  
XX  
XX (CELL-) CELLOMICS INC.  
XX  
XX Guiliano KA, Bright G, Olson K, Burroughs-Tencza S;

XX WPI; 2000-365644/31.  
DR N-PSDB: AAA27573.  
XX  
XX Recombinant nucleic acid encoding a protease biosensor useful for  
PT fluorescence based cell and molecular biochemical assays for drug  
PT discovery comprising three operably linked nucleic acid sequences.  
XX  
XX Claim 14; Fig 29A; 218pp; English.  
XX  
XX The present sequence is that of the EGFP signal domain, which can be  
CC included in novel recombinant protease biosensors (PBs) of the invention.  
CC The PBs (see AAY79638-54) comprise: a first domain (see AAY79579-87)  
CC comprising at least 1 detectable polypeptide signal such as the present  
CC sequence; a second domain (see AAY79588-622) comprising at least 1  
CC protease recognition site; and a third domain (see AAY79623-37)  
CC comprising at least 1 reactant target sequence. A recombinant nucleic  
CC acid (see AAA27627-43) encoding the PB, an expression vector, and a  
CC genetically engineered host cell are also claimed. A claimed method for  
CC identifying compounds that modify protease activity in a cell involves  
CC contacting a host cell that possesses the recombinant PB with a test  
CC compound, and determining the PB distribution in the host cell, where  
CC changes in the distribution of the PB are correlated with modification of  
CC protease activity by the test compound. Claimed kits for identifying  
CC compounds that modify protease activity in a host cell include the  
CC recombinant nucleic acid, or the recombinant PB, or the vector, or the  
CC host cell. The PB is useful in high content screens to detect in vivo  
CC activation of enzymatic activity, and to identify specific activity based  
CC on cleavage of a known recognition motif  
XX  
XX Sequence 239 AA;  
SQ

Query Match 98.7%; Score 1257; DB 3; Length 239;  
Best Local Similarity 98.7%; Pred. No. 2.7e-121;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MYSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTLPFICTTGLKLPVPWPT 60  
Db |||||  
1 MYSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTLPFICTTGLKLPVPWPT 60  
QY 51 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db |||||  
51 LVTLLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
Db |||||  
121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
QY 131 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
Db |||||  
131 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 9  
AAB50804  
ID AAB50804 standard; protein; 239 AA.  
XX  
AC AAB50804;  
XX  
DT 14-MAR-2001 (first entry)  
XX  
DE Jellyfish GFP mutant EGFP.  
XX  
KW Aequorea victoria; jellyfish; fluorescent protein indicator;  
KW green fluorescent protein; GFP; linker moiety; sensor;  
KW calmodulin-binding domain; mutant; mutein.  
XX  
OS Aequorea victoria.  
XX  
PN WO200071565-A2.  
XX  
PD 30-NOV-2000.  
XX

PF 17-MAY-2000; 2000WO-US013684.  
XX  
PR 21-MAY-1999; 99US-00316919.  
PR 21-MAY-1999; 99US-00316920.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
XX  
PI Tsien RY, Baird GA;  
XX  
XX WPI; 2001-032017/04.  
DR N-PSDB; AAC90488.  
XX  
XX Novel fluorescent proteins comprising a sensor protein inserted into  
PT them, useful for measuring the response of a sensor biological, chemical,  
PT electrical or physiological parameter in vivo or in vitro.  
XX  
XX Disclosure; Page 24; 94pp; English.  
XX  
XX The present sequence is a fluorescent protein used in the construction of  
CC a fluorescent protein indicator. The indicator comprises a sensor  
CC polypeptide that is responsive to a chemical, biological, electrical or  
CC physiological parameter, and a fluorescence protein functional group. The  
CC sensor polypeptide is operatively inserted into the fluorescent moiety.  
CC The fluorescent indicator is useful for detecting the presence of a  
CC response inducing member in a sample. The method involves contacting the  
CC sample with the indicator and detecting a change in fluorescence, in  
CC which a change is indicative of the effect of the parameter on the sensor  
CC polypeptide. The novel fluorescent proteins are advantageous due to their  
CC reduced size as compared to the FRET (fluorescence resonance energy  
CC transfer)-based sensors  
XX  
XX Sequence 239 AA;  
SQ

Query Match 98.7%; Score 1257; DB 4; Length 239;  
Best Local Similarity 98.7%; Pred. No. 2.7e-121;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MYSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTLPFICTTGLKLPVPWPT 60  
Db |||||  
1 MYSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTLPFICTTGLKLPVPWPT 60  
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db |||||  
61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
Db |||||  
121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
Db |||||  
181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 10  
AAB85900  
ID AAB85900 standard; protein; 239 AA.  
XX  
AC AAB85900;  
XX  
DT 30-NOV-2001 (first entry)  
XX  
DE A. victoria green fluorescent protein (GFP) and linker sequence.  
XX  
KW Melanin concentrating hormone receptor; MCHR; MCH; chimeric; fusion;  
KW fluorescent polypeptide; orexigenic; anabolic; food intake; GFP;  
KW green fluorescent protein.  
XX  
OS Synthetic.  
OS Aequorea victoria.  
XX  
PN WO200168706-A1.  
XX

PD 20-SEP-2001.  
 XX  
 XX  
 PF 14-MAR-2001; 2001WO-US008071.  
 XX  
 XX  
 PR 15-MAR-2000; 2000US-0189698P.  
 XX  
 XX  
 PA (MERI ) MERCK & CO INC.  
 XX  
 XX  
 XX Marsh DJ;  
 XX  
 XX WPI; 2001-565791/63.  
 DR N-PSDB; AAH47304.  
 XX  
 XX Fusion proteins comprising melanin concentrating hormone receptor  
 PT peptides and fluorescent proteins, useful for identifying appetite  
 PT stimulants.  
 XX  
 XX Claim 2; Page 14; 7lpp; English.  
 XX  
 CC The invention provides melanin concentrating hormone (MCH) receptor  
 CC (MCHR) chimeric and fusion proteins. The MCHR chimeric proteins comprise  
 CC MCHR polypeptide regions from different species. The MCHR fusion protein  
 CC comprise MCHR polypeptide region and a fluorescent polypeptide region  
 CC joined directly, or via a linker, to the carboxy side of the MCHR  
 CC polypeptide region. The MCHR fusion proteins can be expressed by standard  
 CC recombinant methodology. MCH action promotes feeding (orexigenic) and up  
 CC regulation of MCH activity stimulates food intake. The present sequence  
 CC represents a A. victoria green fluorescent protein (GFP) and a linker  
 CC sequence  
 XX  
 XX Sequence 239 AA;  
 SQ

Query Match 98.7%; Score 1257; DB 4; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 2.7e-121;  
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVWPWT 60  
 DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVWPWT 60  
 QY 61 LVTLLSYGQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTLLTYGQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRLEKLGIDFDEGDNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
 DB 121 VNRLEKLGIDFDEGDNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
 QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
 DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 11  
 AAB31171  
 ID AAB31171 standard; protein; 239 AA.  
 XX  
 AC AAB31171;  
 XX  
 XX 02-APR-2001 (first entry)  
 DT  
 XX Amino acid sequence of a green fluorescent protein (GFP).  
 DE  
 XX Growth rate; death rate; reporter gene; luminescent protein;  
 KW fluorescent product; luciferase; green fluorescent protein; GFP.  
 XX  
 OS Aequorea victoria.  
 XX  
 PN WO200075367-A1.  
 XX  
 PD 14-DEC-2000.  
 XX  
 XX 07-JUN-2000; 2000WO-FI000507.  
 PF

XX  
 PR 07-JUN-1999; 99FI-00001296.  
 XX  
 PA (LILI/) LILIUS E.  
 PA (VIRT/) VIRTA M.  
 XX  
 XX Lilius E, Virta M;  
 XX WPI; 2001-061737/07.  
 DR N-PSDB; AAC86954.  
 XX  
 XX Assessing growth and death rates of a micro-organism in a desired  
 PT environment, by introducing 2 reporter genes encoding luminescent and  
 PT fluorescent products and detecting luminescent fluorescence.  
 XX  
 XX Disclosure; Page 27; 32pp; English.  
 XX  
 CC The specification describes a method for assessing the growth rate and  
 CC death rate of a micro-organism within a predetermined time period in a  
 CC desired environment. The method comprises introducing at least two  
 CC reporter genes encoding luminescent and/or fluorescent products into the  
 CC micro-organisms, incubating the micro-organism within the desired  
 CC environment, and detecting luminescence and/or fluorescence after a  
 CC predetermined time period. Use of two different markers within a micro-  
 CC organism enables the differentiation between growth and death rates. The  
 CC method is used to assess the growth rate and death rate of a micro-  
 CC organism within a predetermined time period in a desired environment. The  
 CC present sequence represents a green fluorescent protein (GFP), and is  
 CC encoded by a plasmid which encodes luminescent and fluorescent proteins,  
 CC and is used in the method of the invention  
 XX  
 XX Sequence 239 AA;  
 SQ

Query Match 98.7%; Score 1257; DB 4; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 2.7e-121;  
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVWPWT 60  
 DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVWPWT 60  
 QY 61 LVTLLSYGQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTLLTYGQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRLEKLGIDFDEGDNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
 DB 121 VNRLEKLGIDFDEGDNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
 QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
 DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 12  
 AAG66198  
 ID AAG66198 standard; protein; 239 AA.  
 XX  
 AC AAG66198;  
 XX  
 XX 17-JUN-2002 (first entry)  
 DT  
 XX A. victoria green fluorescent protein (EGFP).  
 DE  
 XX Cyan-green fluorescent protein; fluorescence; recombinant; GFP;  
 KW green fluorescent protein; EGFP.  
 XX  
 OS Aequorea victoria.  
 XX  
 PN JP2002045189-A.  
 XX  
 PD 12-FEB-2002.  
 XX

```
PF 04-AUG-2000; 2000JP-00237165.
XX
PR 04-AUG-2000; 2000JP-00237165.
XX
XX (RIKA ) RIKAGAKU KENKYUSHO.
XX PA
XX WPI: 2002-299190/34.
XX DR N-PSDB; ABL40628.
XX
XX
XX A gene encoding cyan-green fluorescent protein.
XX PT
XX
XX Example; Page 14; 20pp; Japanese.
XX
XX The invention relates to a gene encoding proteins having cyan-green
XX fluorescence characteristic and having a function of showing stable
XX fluorescence characteristic in acid region. A method for the preparation
XX of a cyan-green fluorescent protein is provided which involves a
XX transformant transformed by a recombinant vector comprising the gene,
XX where the transformant is cultured and the protein is collected from the
XX culture. The present sequence represents the A. victoria green
XX fluorescent protein (EGFP)
XX
XX
SQ Sequence 239 AA;
Query Match 98.7%; Score 1257; DB 5; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.7e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKFTCTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKFTCTTGKLPVWPWT 60
QY 51 LVTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNVKTAEVKEFGDTL 120
DB 51 LVTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNVKTAEVKEFGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 131 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKPNKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 131 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKPNKRDHVMVLLGFVTAAGITLGMDELYK 239
RESULT 13
ABG94444
ID ABG94444 standard; protein; 239 AA.
XX
XX ABG94444;
XX
XX 27-NOV-2002 (first entry)
XX
XX Protease biosensor signal sequence #6.
XX
XX Detection; classification; identification; toxin detection; protease;
XX ADP-ribosylating toxin; cytotoxic phospholipase; exfoliative toxin;
XX toxic threat agent.
XX
XX Synthetic.
XX
XX US6416959-B1.
XX
XX 09-JUL-2002.
XX
XX 25-FEB-2000; 2000US-00513783.
XX
XX 27-FEB-1997; 97US-00810983.
XX 27-FEB-1998; 98US-00031271.
XX 26-FEB-1999; 99US-0122152P.
XX 08-MAR-1999; 99US-0123399P.
XX 12-JUL-1999; 99US-00352171.
XX 31-AUG-1999; 99US-0151797P.
XX
PR 17-SEP-1999; 99US-00398965.
PR 29-OCT-1999; 99US-00430656.
PR 01-DEC-1999; 99US-0168408P.
XX
XX (GIUL/) GIULIANO K.
XX PA (KAPU/) KAPUR R.
XX
XX Giuliano K, Kapur R;
XX
XX WPI: 2002-634730/68.
XX DR N-PSDB; ABS71491.
XX
XX Automated cell-based toxin detection, classification, and/or
XX identification by treating cells involves use of three classes of
XX luminescent reporter molecules such as detectors, classifiers or
XX identifiers.
XX
XX Example 10; Fig 29A; 214pp; English.
XX
XX The invention describes methods of automated detection, classification
XX and identification comprising treating cells containing luminescent
XX reporter molecules (i) in array of locations with a test substance, where
XX (i) are detectors, classifiers or identifiers, imaging cells in each
XX location to obtain luminescent signals and converting optical information
XX into digital data to interpret presence of toxins in the test substance.
XX The method are useful for detection of toxins chosen from proteases, ADP-
XX ribosylating toxins, cytotoxic phospholipases, and exfoliative toxins.
XX Three classes of cell-based luminescent reporter molecules such as
XX detectors, classifiers and identifiers are described and serve as
XX reporters of toxic threat agents. The first two levels of
XX characterisation ensure a rapid readout of toxin class without
XX sacrificing the ability to detect many new mutant toxins or dissect
XX several complex mixtures of known toxins. This is the amino acid sequence
XX of a protease biosensor related signal sequence used in the cell-based
XX screening system
XX
XX
SQ Sequence 239 AA;
Query Match 98.7%; Score 1257; DB 5; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.7e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKFTCTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKFTCTTGKLPVWPWT 60
QY 51 LVTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNVKTAEVKEFGDTL 120
DB 51 LVTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNVKTAEVKEFGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 131 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKPNKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 131 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKPNKRDHVMVLLGFVTAAGITLGMDELYK 239
RESULT 14
AAE14599
ID AAE14599 standard; protein; 239 AA.
XX
XX AAE14599;
XX
XX 31-MAY-2002 (first entry)
XX
XX Aequorea victoria enhanced green fluorescent protein.
XX DE
XX Mutagenesis; enhanced green fluorescent protein; EGFP; mutant; muten.
XX
XX Aequorea victoria.
XX OS Synthetic.
```

XX Key Location/Qualifiers  
 FT Misc-difference 1. .3 /note= "Wild-type GFP Met-Ser are replaced with Met-Val-Ser"  
 FT Misc-difference 65 /note= "GFP Phe64 is replaced by Leu"  
 FT Misc-difference 66 /note= "GFP Ser65 is replaced by Thr"  
 XX EP1178109-A1.  
 XX 06-FEB-2002.  
 XX 03-AUG-2001; 2001EP-00306650.  
 XX 04-AUG-2000; 2000JP-00237166.  
 XX (RIKE ) RIKEN KK.  
 XX Miyawaki A, Sawano A;  
 PI WPI; 2002-208112/27.  
 DR N-PSDB; AAD27910.  
 XX Method for mutagenesis, e.g. for introducing certain or random mutations at certain sites of the nucleotide sequence, comprises synthesizing a mutated strand and a complementary strand by use of megaprimers.  
 PS Example 1; Page 13-14; 31pp; English.  
 XX The invention relates to a method for mutagenesis that comprises synthesizing a mutated strand and a complementary strand by use of megaprimers. The method basically comprises a DNA synthesis in which one or more primers that have a nucleotide sequence containing at least one mutation and a phosphorylated 5'-terminus are annealed to a template DNA and then subjected to an elongation reaction using a thermostable high-fidelity DNA polymerase, after which the phosphorylated 5'-terminus and the elongated terminus are ligated by means of a thermostable DNA ligase to synthesise a circular DNA containing the primers; a digestion in which the step of DNA synthesis is repeated several times to amplify the DNA containing the primers and then, at least DNAs other than the amplified circular DNA are digested into several fragments; and a double-stranded DNA synthesis in which, with the several fragments obtained in the step of digestion as megaprimers, the megaprimers are annealed to the circular DNA synthesised above, followed by an elongation reaction performed using the thermostable high-fidelity DNA polymerase. The method is useful for mutagenesis, particularly for introducing certain mutations at certain sites of a nucleotide sequence, or for introducing random mutations at certain sites of the nucleotide sequence. The present method is simple, speedy, economical and widely applicable. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used for mutagenesis in an exemplification of the invention. The EGFP is derived by altering the green fluorescent protein (GFP) sequence of Aequorea victoria  
 XX Sequence 239 AA;  
 Query Match 98.7%; Score 1257; DB 5; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 2.7e-121;  
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MYSKGELFTGVVPIVLVDGVDNGHKFVSVEGEGDATYGKLTAKFICTTGKLPVPWPT 60  
 DB 1 MYSKGELFTGVVPIVLVDGVDNGHKFVSVEGEGDATYGKLTAKFICTTGKLPVPWPT 60  
 QY 61 LVTLLSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120  
 DB 61 LVTLLTYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGDIDFKEDGNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 DB 121 VNRIELKGDIDFKEDGNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
 DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
 RESULT 15  
 AAE34958  
 ID AAE34958 standard; protein; 239 AA.  
 XX AAE34958;  
 AC AAE34958;  
 XX 28-MAY-2003 (first entry)  
 XX Aequorea victoria enhanced green fluorescent protein (EGFP).  
 XX Phosphorylation indicator; fluorescent protein; detection; phosphatase;  
 KW kinase; enhanced green fluorescent protein; EGFP.  
 XX Aequorea victoria.  
 OS WO200295058-A2.  
 XX 28-NOV-2002.  
 XX 24-MAY-2002; 2002WO-US016955.  
 XX 24-MAY-2001; 2001US-00865291.  
 XX (REGC ) UNIV CALIFORNIA.  
 XX Tsien RY, Ting AY, Zhang J;  
 WPI: 2003-148474/14.  
 DR N-PSDB; AAD53428.  
 XX Novel chimeric phosphorylation indicators, useful for detecting kinase/phosphatase in samples, has donor molecule, phosphorylatable domain, phosphaminoacid binding domain, and acceptor molecule, in operative linkage.  
 PS Disclosure; Col 56-57; 38pp; English.  
 XX The present invention relates to chimeric phosphorylation indicators comprising a phosphorylation polypeptide and a fluorescent protein or in operative linkage, a donor molecule, a phosphorylatable domain, a phosphaminoacid binding domain (PAABD) and an acceptor molecule. The phosphorylation indicators of the invention are useful for detecting kinases or phosphatases in a biological sample. They are also useful in high throughput analysis e.g. for detecting a kinase inhibitor or phosphatase inhibitor. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used in the invention  
 XX Sequence 239 AA;  
 Query Match 98.7%; Score 1257; DB 6; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 2.7e-121;  
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MYSKGELFTGVVPIVLVDGVDNGHKFVSVEGEGDATYGKLTAKFICTTGKLPVPWPT 60  
 DB 1 MYSKGELFTGVVPIVLVDGVDNGHKFVSVEGEGDATYGKLTAKFICTTGKLPVPWPT 60  
 QY 61 LVTLLSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120  
 DB 61 LVTLLTYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGDIDFKEDGNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 DB 121 VNRIELKGDIDFKEDGNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
 DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

us-09-887-784-641.rag

Sun Jun 27 18:27:37 2004

Db 131 DHYQONTPIGCGFVLLPDNHYLSTOSALS KDPNEKRDHNVLLFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 15:54:10  
Job time : 48.1111 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:55 ; Search time 12.7778 Seconds  
(without alignments)  
965.630 Million cell updates/sec

Title: US-09-887-784-64L  
Perfect score: 1273  
Sequence: 1 MVSKEELFTGVVPIILVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1257	98.7	239	3	US-09-172-063-3
2	1257	98.7	239	4	US-09-513-783A-46
3	1257	98.7	239	4	US-09-316-919-4
4	1257	98.7	239	4	US-09-602-641-3
5	1257	98.7	239	4	US-09-920-922-2
6	1257	98.7	281	3	US-09-062-102-1
7	1257	98.7	281	4	US-09-364-946-1
8	1257	98.7	294	4	US-09-513-783A-2
9	1257	98.7	323	3	US-09-172-063-21
10	1257	98.7	323	4	US-09-602-641-21
11	1257	98.7	364	3	US-09-085-305-6
12	1257	98.7	379	4	US-09-417-197-129
13	1257	98.7	434	4	US-09-800-170-48
14	1257	98.7	442	4	US-09-417-197-127
15	1257	98.7	459	4	US-09-513-783A-170
16	1257	98.7	544	4	US-09-417-197-113
17	1257	98.7	544	4	US-09-417-197-115
18	1257	98.7	604	4	US-09-417-197-59
19	1257	98.7	605	4	US-09-417-197-41
20	1257	98.7	606	4	US-09-417-197-65
21	1257	98.7	607	4	US-09-417-197-47
22	1257	98.7	630	4	US-09-417-197-63
23	1257	98.7	631	4	US-09-417-197-39
24	1257	98.7	633	4	US-09-417-197-45
25	1257	98.7	635	4	US-09-417-197-125
26	1257	98.7	642	2	US-08-818-253-2
27	1257	98.7	642	2	US-08-818-253-6

28	1257	98.7	642	3	US-08-818-252-2	Sequence 2, Appli
29	1257	98.7	642	3	US-08-818-252-6	Sequence 6, Appli
30	1257	98.7	652	3	US-08-818-253-4	Sequence 4, Appli
31	1257	98.7	652	3	US-08-818-252-4	Sequence 4, Appli
32	1257	98.7	718	4	US-09-417-197-75	Sequence 75, Appli
33	1257	98.7	719	4	US-09-417-197-51	Sequence 51, Appli
34	1257	98.7	726	4	US-09-417-197-71	Sequence 71, Appli
35	1257	98.7	727	4	US-09-417-197-139	Sequence 139, App
36	1257	98.7	783	4	US-09-513-783A-176	Sequence 176, App
37	1257	98.7	797	4	US-09-417-197-141	Sequence 141, App
38	1257	98.7	797	4	US-09-417-197-143	Sequence 143, App
39	1257	98.7	798	4	US-09-417-197-77	Sequence 77, Appli
40	1257	98.7	805	4	US-09-513-783A-178	Sequence 178, App
41	1257	98.7	806	4	US-09-417-197-53	Sequence 53, Appli
42	1257	98.7	836	4	US-09-417-197-61	Sequence 61, Appli
43	1257	98.7	842	4	US-09-417-197-43	Sequence 43, Appli
44	1257	98.7	843	4	US-09-417-197-117	Sequence 117, App
45	1257	98.7	853	4	US-09-417-197-119	Sequence 119, App

ALIGNMENTS

RESULT 1  
US-09-172-063-3  
; Sequence 3, Application US/09172063  
; Patent No. 6150176  
; GENERAL INFORMATION:  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Miyawaki, Atsushi  
; APPLICANT: Liopis, Juan  
; APPLICANT: Wachter, Rebekka M.  
; APPLICANT: Remington, S. James  
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
; FILE REFERENCE: 07257/071001  
; CURRENT APPLICATION NUMBER: US/09/172,063  
; CURRENT FILING DATE: 1998-10-13  
; EARLIER APPLICATION NUMBER: 09/094,359  
; EARLIER FILING DATE: 1998-06-09  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aquorea victoria  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: EGFP  
US-09-172-063-3

QY	1	MVSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGTGKLPVDPWT	60	Query Match	98.7%;	Score 1257;	DB 3;	Length 239;
Db	1	MVSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGTGKLPVDPWT	60	Best Local Similarity	98.7%;	Pred. No. 9.3e-127;		
QY	61	LVTLLSYGVCFSRYPDHMKQHDFFKSAPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL	120	Matches 236;	Conservative	1;	Mismatches	2;
Db	61	LVTLLTYGVCFSRYPDHMKQHDFFKSAPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL	120				Indels	0;
QY	121	VNRIELKGDIFDKEDGNILGHKLEYNYSNHYIMADQKNIGKVNFKIRHNIEDGSVOLA	180					
Db	121	VNRIELKGDIFDKEDGNILGHKLEYNYSNHYIMADQKNIGKVNFKIRHNIEDGSVOLA	180					
QY	181	DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITTLGMDELYK	239					
Db	181	DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITTLGMDELYK	239					

Qy 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 Db 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLGFVTAAGITLGMDELYK 239  
 Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 4  
 US-09-602-641-3  
 ; Sequence 3, Application US/09602641  
 ; Patent No. 6608189  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tsien, Roger Y.  
 ; APPLICANT: Miyawaki, Atsushi  
 ; APPLICANT: Llopis, Juan  
 ; APPLICANT: Wachter, Rebekka M.  
 ; APPLICANT: Remington, S. James  
 ; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
 ; MEASURING THE PH OF A BIOLOGICAL SAMPLE  
 ; FILE REFERENCE: 07257/071001  
 ; CURRENT APPLICATION NUMBER: US/09/602,641  
 ; CURRENT FILING DATE: 2000-06-22  
 ; PRIOR APPLICATION NUMBER: 09/172,063  
 ; PRIOR FILING DATE: 1998-10-13  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 3  
 ; LENGTH: 239  
 ; TYPE: PRT  
 ; ORGANISM: Aequorea victoria  
 ; FEATURE:  
 ; NAME/KEY: VARIANT  
 ; LOCATION: (0)...(0)  
 ; OTHER INFORMATION: EGFP  
 US-09-602-641-3

Query Match 98.7%; Score 1257; DB 4; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 9.3e-127;  
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MVSKEELFTGVVPIVLVDGVDVNGHFKFSVSGEGDATYKLTLPFICTTGKLPVPWPT 60  
 Db 1 MVSKEELFTGVVPIVLVDGVDVNGHFKFSVSGEGDATYKLTLPFICTTGKLPVPWPT 60  
 Qy 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 Db 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 Qy 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 Db 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLGFVTAAGITLGMDELYK 239  
 Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 5  
 US-09-920-922-2  
 ; Sequence 2, Application US/09920922  
 ; Patent No. 6673610  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Miyawaki, Atsushi  
 ; APPLICANT: Sawano, Asako  
 ; TITLE OF INVENTION: METHOD FOR MUTAGENESIS  
 ; FILE REFERENCE: 11283-012001  
 ; CURRENT APPLICATION NUMBER: US/09/920,922  
 ; CURRENT FILING DATE: 2001-08-02  
 ; PRIOR APPLICATION NUMBER: JP 2000-237166  
 ; PRIOR FILING DATE: 2000-08-04  
 ; NUMBER OF SEQ ID NOS: 9

RESULT 2  
 US-09-513-783A-46  
 ; Sequence 46, Application US/09513783A  
 ; Patent No. 6416959  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Giuliano, Kenneth A.  
 ; APPLICANT: Kapur, Ravi  
 ; TITLE OF INVENTION: A System for Cell Based Screening  
 ; FILE REFERENCE: 97-022-L1  
 ; CURRENT APPLICATION NUMBER: US/09/513,783A  
 ; CURRENT FILING DATE: 2000-02-25  
 ; NUMBER OF SEQ ID NOS: 180  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 46  
 ; LENGTH: 239  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: EGFP  
 US-09-513-783A-46

Query Match 98.7%; Score 1257; DB 4; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 9.3e-127;  
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MVSKEELFTGVVPIVLVDGVDVNGHFKFSVSGEGDATYKLTLPFICTTGKLPVPWPT 60  
 Db 1 MVSKEELFTGVVPIVLVDGVDVNGHFKFSVSGEGDATYKLTLPFICTTGKLPVPWPT 60  
 Qy 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 Db 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 Qy 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 Db 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLGFVTAAGITLGMDELYK 239  
 Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 3  
 US-09-316-919-4  
 ; Sequence 4, Application US/09316919  
 ; Patent No. 6469154  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tsien, Roger Y.  
 ; APPLICANT: Baird, Geoffrey  
 ; TITLE OF INVENTION: FLUORESCENT PROTEIN INDICATORS  
 ; FILE REFERENCE: 07257/073001  
 ; CURRENT APPLICATION NUMBER: US/09/316,919  
 ; CURRENT FILING DATE: 1999-05-21  
 ; NUMBER OF SEQ ID NOS: 63  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 239  
 ; TYPE: PRT  
 ; ORGANISM: Aequorea victoria  
 US-09-316-919-4

Query Match 98.7%; Score 1257; DB 4; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 9.3e-127;  
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MVSKEELFTGVVPIVLVDGVDVNGHFKFSVSGEGDATYKLTLPFICTTGKLPVPWPT 60  
 Db 1 MVSKEELFTGVVPIVLVDGVDVNGHFKFSVSGEGDATYKLTLPFICTTGKLPVPWPT 60  
 Qy 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 Db 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120



```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match          98.7%; Score 1257; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 9.3e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKFSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKFSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLCFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLCFVTAAGITLGMDELYK 239

RESULT 6
US-09-062-102-1
; Sequence 1, Application US/09062102
; Patent No. 6130313
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100
; CURRENT APPLICATION NUMBER: US/09/062,102
; CURRENT FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: US 60/060,855
; EARLIER FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6130313
US-09-062-102-1

Query Match          98.7%; Score 1257; DB 3; Length 281;
Best Local Similarity 98.7%; Pred. No. 1.2e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKFSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKFSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLCFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLCFVTAAGITLGMDELYK 239

RESULT 7
US-09-364-946-1
; Sequence 1, Application US/09364946
; Patent No. 6306600
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100CIP/D2
; CURRENT APPLICATION NUMBER: US/09/364,946
; CURRENT FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: US 09/191,233
; EARLIER FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6306600
US-09-364-946-1

Query Match          98.7%; Score 1257; DB 4; Length 281;
Best Local Similarity 98.7%; Pred. No. 1.2e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKFSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKFSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLCFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLCFVTAAGITLGMDELYK 239

RESULT 8
US-09-513-783A-2
; Sequence 2, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GFP-DEVD-Annexin II construct
US-09-513-783A-2

Query Match          98.7%; Score 1257; DB 4; Length 294;
Best Local Similarity 98.7%; Pred. No. 1.3e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKFSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKFSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLCFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLCFVTAAGITLGMDELYK 239
```



ATTORNEY/AGENT INFORMATION:  
NAME: Francis, Carol L  
REGISTRATION NUMBER: 36,513  
REFERENCE/DOCKET NUMBER: 6510/102US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-327-3400  
TELEFAX: 650-327-3231  
TELEX:  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 364 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-085-305-6

Query Match 98.7%; Score 1257; DB 3; Length 364;  
Best Local Similarity 98.7%; Pred. No. 1.9e-126;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60  
DB 126 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 185

QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 186 LVTLLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 245

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 246 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 305

QY 181 DHYQONTPIGDGVVLPDNNHLSQTSALSQSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB 306 DHYQONTPIGDGVVLPDNNHLSQTSALSQSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 364

RESULT 12  
US-09-417-197-129  
Sequence 129, Application US/09417197  
Patent No. 6518021  
GENERAL INFORMATION:  
APPLICANT: Ole THASTRUP, et al.  
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An  
FILE OF INVENTION: On A Cellular Response  
FILE REFERENCE: 3759-0110P  
CURRENT APPLICATION NUMBER: US/09/417,197  
CURRENT FILING DATE: 1999-10-07  
NUMBER OF SEQ ID NOS: 143  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 129  
LENGTH: 379  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: actin-binding-domain-EGFP fusion  
US-09-417-197-129

Query Match 98.7%; Score 1257; DB 4; Length 379;  
Best Local Similarity 98.7%; Pred. No. 1.9e-126;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60  
DB 141 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 200

QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 201 LVTLLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 260

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 261 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 320

QY 181 DHYQONTPIGDGVVLPDNNHLSQTSALSQSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB 321 DHYQONTPIGDGVVLPDNNHLSQTSALSQSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 379

RESULT 13  
US-09-800-170-48  
Sequence 48, Application US/09800170  
Patent No. 6481667  
GENERAL INFORMATION:  
APPLICANT: Kinsella, Todd  
TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES  
FILE REFERENCE: A-68614-1/DJB/RMS/RMK  
CURRENT APPLICATION NUMBER: US/09/800,170  
CURRENT FILING DATE: 2001-08-28  
PRIOR APPLICATION NUMBER: US 60/187,130  
PRIOR FILING DATE: 2000-03-06  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 48  
LENGTH: 434  
TYPE: PRT  
ORGANISM: Synechocystis PCC6803  
US-09-800-170-48

Query Match 98.7%; Score 1257; DB 4; Length 434;  
Best Local Similarity 98.7%; Pred. No. 2.3e-126;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60  
DB 196 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 255

QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 256 LVTLLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 315

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 316 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 375

QY 181 DHYQONTPIGDGVVLPDNNHLSQTSALSQSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB 376 DHYQONTPIGDGVVLPDNNHLSQTSALSQSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 434

RESULT 14  
US-09-417-197-127  
Sequence 127, Application US/09417197  
Patent No. 6518021  
GENERAL INFORMATION:  
APPLICANT: Ole THASTRUP, et al.  
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An  
FILE OF INVENTION: On A Cellular Response  
FILE REFERENCE: 3759-0110P  
CURRENT APPLICATION NUMBER: US/09/417,197  
CURRENT FILING DATE: 1999-10-07  
NUMBER OF SEQ ID NOS: 143  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 127  
LENGTH: 442  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: EGFP-RhoA fusion  
US-09-417-197-127

Query Match 98.7%; Score 1257; DB 4; Length 442;  
Best Local Similarity 98.7%; Pred. No. 2.3e-126;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60

Db	1	MVSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGGDATYGKLTLCFICTTGKLPVPWPT	60
Qy	61	LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL	120
Db	61	LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL	120
Qy	121	VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA	180
Db	121	VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA	180
Qy	181	DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK	239
Db	181	DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK	239

RESULT 15  
US-09-513-783A-170  
; Sequence 170, Application US/09513783A  
; Patent No. 641659  
; GENERAL INFORMATION:  
; APPLICANT: Giuliano, Kenneth A.  
; APPLICANT: Kapur, Ravi  
; TITLE OF INVENTION: A System for Cell Based Screening  
; FILE REFERENCE: 97-022-L1  
; CURRENT APPLICATION NUMBER: US/09/513,783A  
; CURRENT FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 170  
; LENGTH: 459  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP27  
US-09-513-783A-170

Query Match	98.7%	Score 1257;	DB 4;	Length 459;
Best Local Similarity	98.7%	Pred. No. 2.5e-126;		
Matches	236;	Conservative	1;	Mismatches 2; Indels 0; Gaps 0;
Qy	1	MVSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGGDATYGKLTLCFICTTGKLPVPWPT	60	
Db	1	MVSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGGDATYGKLTLCFICTTGKLPVPWPT	60	
Qy	61	LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL	120	
Db	61	LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL	120	
Qy	121	VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA	180	
Db	121	VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA	180	
Qy	181	DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK	239	
Db	181	DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK	239	

Search completed: June 21, 2004, 16:04:02  
Job time : 13.778 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: June 21, 2004, 15:54:20 ; Search time 34.7778 Seconds  
(without alignments)  
1940.117 Million cell updates/sec

Title: US-09-887-784-641

Perfect score: 1273

Sequence: 1 MVSKEELFTGVVPIVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1268	99.6	239	9 US-09-887-784-4	Sequence 4, Appli
2	1268	99.6	239	12 US-10-296-953-4	Sequence 4, Appli
3	1268	99.6	239	14 US-10-270-223-6	Sequence 6, Appli
4	1268	99.6	893	14 US-10-257-909A-30	Sequence 30, Appl
5	1268	99.6	1132	14 US-10-257-909A-32	Sequence 32, Appl
6	1260	99.0	239	9 US-09-887-784-2	Sequence 2, Appli
7	1260	99.0	239	12 US-10-296-953-2	Sequence 2, Appli
8	1257	98.7	239	9 US-09-920-922-2	Sequence 2, Appli
9	1257	98.7	239	9 US-09-999-745-4	Sequence 4, Appli
10	1257	98.7	239	10 US-09-866-538-4	Sequence 4, Appli
11	1257	98.7	239	10 US-09-797-496B-2	Sequence 2, Appli
12	1257	98.7	239	10 US-09-794-308-4	Sequence 4, Appli
13	1257	98.7	239	10 US-09-865-291-4	Sequence 4, Appli
14	1257	98.7	239	12 US-10-457-982-3	Sequence 3, Appli
15	1257	98.7	239	14 US-10-121-258-13	Sequence 13, Appl

16	1257	98.7	239	14	US-10-221-461-7	Sequence 7, Appli
17	1257	98.7	239	14	US-10-100-957A-46	Sequence 46, Appl
18	1257	98.7	239	14	US-10-177-390-2	Sequence 2, Appli
19	1257	98.7	239	14	US-10-338-411-3	Sequence 3, Appli
20	1257	98.7	239	15	US-10-370-570-4	Sequence 4, Appli
21	1257	98.7	239	15	US-10-389-640-3	Sequence 3, Appli
22	1257	98.7	259	14	US-10-314-861-11	Sequence 11, Appl
23	1257	98.7	281	12	US-09-931-232-1	Sequence 1, Appli
24	1257	98.7	288	14	US-10-314-861-37	Sequence 37, Appl
25	1257	98.7	293	14	US-10-314-861-35	Sequence 35, Appl
26	1257	98.7	294	14	US-10-100-957A-2	Sequence 2, Appli
27	1257	98.7	295	14	US-10-314-861-39	Sequence 39, Appl
28	1257	98.7	299	14	US-10-314-861-33	Sequence 33, Appl
29	1257	98.7	305	14	US-10-314-861-31	Sequence 31, Appl
30	1257	98.7	308	14	US-10-033-717-35	Sequence 35, Appl
31	1257	98.7	311	14	US-10-314-861-29	Sequence 29, Appl
32	1257	98.7	320	14	US-10-338-411-11	Sequence 11, Appl
33	1257	98.7	320	15	US-10-389-640-11	Sequence 11, Appl
34	1257	98.7	323	12	US-10-457-982-21	Sequence 21, Appl
35	1257	98.7	323	14	US-10-338-411-7	Sequence 7, Appli
36	1257	98.7	323	14	US-10-338-411-13	Sequence 13, Appl
37	1257	98.7	323	15	US-10-389-640-7	Sequence 7, Appli
38	1257	98.7	323	15	US-10-389-640-13	Sequence 13, Appl
39	1257	98.7	324	14	US-10-314-861-16	Sequence 16, Appl
40	1257	98.7	345	14	US-10-338-411-5	Sequence 5, Appli
41	1257	98.7	345	15	US-10-389-640-5	Sequence 5, Appli
42	1257	98.7	346	14	US-10-338-411-9	Sequence 9, Appli
43	1257	98.7	346	15	US-10-389-640-9	Sequence 9, Appli
44	1257	98.7	359	14	US-10-033-717-33	Sequence 33, Appl
45	1257	98.7	359	14	US-10-033-717-34	Sequence 34, Appl

## ALIGNMENTS

## RESULT 1

US-09-887-784-4  
; Sequence 4, Application US/09887784  
; Patent No. US20020177189A1  
; GENERAL INFORMATION:  
; APPLICANT: BJORN, Sara et al  
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS  
; FILE REFERENCE: 3759-0115P  
; CURRENT APPLICATION NUMBER: US/09/887,784  
; CURRENT FILING DATE: 2001-06-19  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 4  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequoria Victoria  
US-09-887-784-4

Query Match 99.6%; Score 1268; DB 9; Length 239;

Best Local Similarity 99.6%; Pred. No. 4e-124;

Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MVSKEELFTGVVPIVELDGVNGHKFSVSGEGDATYTKLTKFKICTTGKLPVWPPT 60
Db	1	MVSKEELFTGVVPIVELDGVNGHKFSVSGEGDATYTKLTKFKICTTGKLPVWPPT 60
Qy	61	LVTLLSYGVQCFSRYPDHMKQHDFFKSAWPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db	61	LVTLLSYGVQCFSRYPDHMKQHDFFKSAWPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Qy	121	VNRIELKIDFDKEDGNILGHKLEYNHNHYNIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db	121	VNRIELKIDFDKEDGNILGHKLEYNHNHYNIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Qy	181	DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
Db	181	DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

```
RESULT 2
US-10-296-953-4
; Sequence 4, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PLO095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-4

Query Match          99.6%; Score 1268; DB 12; Length 239;
Best Local Similarity 99.6%; Pred. No. 4e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MWSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFKICTTGLKLPVPWPT 60
DB 1 MWSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFKICTTGLKLPVPWPT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 3
US-10-270-223-6
; Sequence 6, Application US/10270223
; Publication No. US20030143634A1
; GENERAL INFORMATION:
; APPLICANT: Bioimage A/S
; TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR COMPONENTS
; TITLE OF INVENTION: INTERACT LIVING CELLS, AND TO EXTRACT QUANTITATIVE INFORMATION FROM INTERACTIONS BY FLUORESCENCE REDISTRIBUTION.
; FILE REFERENCE: 3759-0126P
; CURRENT APPLICATION NUMBER: US/10/270,223
; CURRENT FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 6
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Aequorea Victoria and Human
US-10-270-223-6

Query Match          99.6%; Score 1268; DB 14; Length 363;
Best Local Similarity 99.6%; Pred. No. 7.4e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MWSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFKICTTGLKLPVPWPT 60
DB 1 MWSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFKICTTGLKLPVPWPT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 4
US-10-257-909A-30
; Sequence 30, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 893
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-30

Query Match          99.6%; Score 1268; DB 14; Length 893;
Best Local Similarity 99.6%; Pred. No. 2.7e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MWSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFKICTTGLKLPVPWPT 60
DB 655 MWSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFKICTTGLKLPVPWPT 714
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 715 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 774
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 775 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 834
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 835 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLLGFVTAAGITLGMDELYK 893

RESULT 5
US-10-257-909A-32
; Sequence 32, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
QY 1 MWSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFKICTTGLKLPVPWPT 60
DB 1 MWSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFKICTTGLKLPVPWPT 60
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 4
US-10-257-909A-30
; Sequence 30, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 893
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-30

Query Match          99.6%; Score 1268; DB 14; Length 893;
Best Local Similarity 99.6%; Pred. No. 2.7e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MWSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFKICTTGLKLPVPWPT 60
DB 655 MWSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFKICTTGLKLPVPWPT 714
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 715 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 774
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 775 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 834
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 835 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLLGFVTAAGITLGMDELYK 893

RESULT 5
US-10-257-909A-32
; Sequence 32, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
```

```

; SEQ ID NO 32
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-32

Query Match          99.6%; Score 1268; DB 14; Length 1132;
Best Local Similarity 99.6%; Pred. No. 3.8e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
Db 894 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 953

QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 954 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 1013

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 1014 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 1073

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 1074 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 1132

RESULT 6
US-09-887-784-2
; Sequence 2, Application US/09887784
; Patent No. US20020177189A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea Victoria
US-09-887-784-2

Query Match          99.0%; Score 1260; DB 9; Length 239;
Best Local Similarity 99.2%; Pred. No. 2.8e-123;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60

QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 7
US-10-296-953-2
; Sequence 2, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match          98.7%; Score 1257; DB 9; Length 239;
Best Local Similarity 98.7%; Pred. No. 5.7e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
```

```

; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match          99.0%; Score 1260; DB 12; Length 239;
Best Local Similarity 99.2%; Pred. No. 2.8e-123;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60

QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 8
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. US20020083488A1
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match          98.7%; Score 1257; DB 9; Length 239;
Best Local Similarity 98.7%; Pred. No. 5.7e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
```

QY 61 LVTLLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTLLTYGVQCFSPYDPHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQOQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQOQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239  
RESULT 9  
US-09-999-745-4  
; Sequence 4, Application US/09999745  
; Patent No. US20020157120A1  
; GENERAL INFORMATION:  
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Baird, Geoffrey  
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS  
; FILE REFERENCE: REGEN1470-1  
; CURRENT APPLICATION NUMBER: US/09/999,745  
; PRIOR FILING DATE: 2001-10-23  
; PRIOR FILING DATE: 1999-05-21  
; NUMBER OF SEQ ID NOS: 67  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
US-09-999-745-4

Query Match 98.7%; Score 1257; DB 9; Length 239;  
Best Local Similarity 98.7%; Pred. No. 5.7e-123;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MVSKEBELFTGVVPIILVELDGDVNGHKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60  
DB 1 MVSKEBELFTGVVPIILVELDGDVNGHKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60  
QY 51 LVTLLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 51 LVTLLTYGVQCFSPYDPHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 131 DHYQOQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239  
DB 131 DHYQOQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

RESULT 10  
US-09-866-538-4  
; Sequence 4, Application US/09866538  
; Publication No. US2003032088A1  
; GENERAL INFORMATION:  
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA  
; APPLICANT: TSIEEN, Roger  
; APPLICANT: Campbell, Robert  
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS  
; FILE REFERENCE: REGEN1530-2  
; CURRENT APPLICATION NUMBER: US/09/866,538  
; PRIOR FILING DATE: 2001-05-24  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 239  
; TYPE: PRT

; ORGANISM: Aequorea victoria  
US-09-866-538-4  
Query Match 98.7%; Score 1257; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 5.7e-123;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MVSKEBELFTGVVPIILVELDGDVNGHKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60  
DB 1 MVSKEBELFTGVVPIILVELDGDVNGHKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60  
QY 61 LVTLLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTLLTYGVQCFSPYDPHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQOQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQOQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239  
RESULT 11  
US-09-797-496B-2  
; Sequence 2, Application US/09797496B  
; Publication No. US20030049597A1  
; GENERAL INFORMATION:  
; APPLICANT: Simon, Sanford M.  
; APPLICANT: Chen, Yu  
; TITLE OF INVENTION: Chimeric Fluorescent Enzymes and Uses Thereof  
; FILE REFERENCE: 600-1-267  
; CURRENT APPLICATION NUMBER: US/09/797,496B  
; CURRENT FILING DATE: 2002-05-24  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Aequorea victoria green fluorescent protein modified as described  
; OTHER INFORMATION: in specification  
US-09-797-496B-2

Query Match 98.7%; Score 1257; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 5.7e-123;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MVSKEBELFTGVVPIILVELDGDVNGHKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60  
DB 1 MVSKEBELFTGVVPIILVELDGDVNGHKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60  
QY 61 LVTLLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTLLTYGVQCFSPYDPHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQOQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQOQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239  
RESULT 12  
US-09-794-308-4  
; Sequence 4, Application US/09794308  
; Publication No. US20030170911A1  
; GENERAL INFORMATION:  
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA



APPLICANT: TSUEN, Roger  
APPLICANT: ZACHARIAS, David  
APPLICANT: BAIRD, Geoffrey  
TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS  
FILE REFERENCE: REG1530  
CURRENT APPLICATION NUMBER: US/09/794,308  
CURRENT FILING DATE: 2001-02-26  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
US-09-794-308-4

Query Match 98.7%; Score 1257; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 5.7e-123;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQKPNKRDHMLVLCFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQKPNKRDHMLVLCFVTAAGITLGMDELYK 239

RESULT 13  
US-09-865-291-4  
Sequence 4, Application US/09865291  
Publication No. US20030186229A1  
GENERAL INFORMATION:  
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA  
APPLICANT: TSUEN, Roger  
APPLICANT: TING, Alice  
APPLICANT: ZHANG, Jin  
TITLE OF INVENTION: EMISSION RATIONETRIC INDICATORS OF PHOSPHORYLATION  
FILE REFERENCE: REG1550  
CURRENT APPLICATION NUMBER: US/09/865,291  
CURRENT FILING DATE: 2001-05-24  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
US-09-865-291-4

Query Match 98.7%; Score 1257; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 5.7e-123;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180

QY 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQKPNKRDHMLVLCFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQKPNKRDHMLVLCFVTAAGITLGMDELYK 239

RESULT 14  
US-10-457-982-3  
Sequence 3, Application US/10457982  
Publication No. US20030212265A1  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger Y.  
APPLICANT: Miyawaki, Atsushi  
APPLICANT: Liopis, Juan  
APPLICANT: Wachter, Rebekka M.  
APPLICANT: Remington, S. James  
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
FILE REFERENCE: 07257/071001  
CURRENT APPLICATION NUMBER: US/10/457,982  
CURRENT FILING DATE: 2003-06-09  
PRIOR APPLICATION NUMBER: US/09/602,641  
PRIOR FILING DATE: 2000-06-22  
PRIOR APPLICATION NUMBER: 09/172,063  
PRIOR FILING DATE: 1998-10-13  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (0)...(0)  
OTHER INFORMATION: EGFP  
US-10-457-982-3

Query Match 98.7%; Score 1257; DB 12; Length 239;  
Best Local Similarity 98.7%; Pred. No. 5.7e-123;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQKPNKRDHMLVLCFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQKPNKRDHMLVLCFVTAAGITLGMDELYK 239

RESULT 15  
US-10-121-258-13  
Sequence 13, Application US/10121258  
Publication No. US20030059835A1  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger  
APPLICANT: Campbell, Robert  
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT  
TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME  
FILE REFERENCE: UC083.1CP2CP1  
CURRENT APPLICATION NUMBER: US/10/121,258  
CURRENT FILING DATE: 2002-04-10  
PRIOR APPLICATION NUMBER: 09/794,308  
PRIOR FILING DATE: 2001-02-26  
PRIOR APPLICATION NUMBER: 09/866,538  
PRIOR FILING DATE: 2001-05-24

```
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Enhanced Green Fluorescent Protein (EGFP)
US-10-121-258-13

Query Match      98.7%; Score 1257; DB 14; Length 239;
Best Local Similarity 98.7%; Pred. No. 5.7e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1  MWSKGEELFTGVVPIILVELDGDVNGHKFVSVSSEGDATYVKLTLPFICTTGKLPVPWPT 60
Db      1  MWSKGEELFTGVVPIILVELDGDVNGHKFVSVSSEGDATYVKLTLPFICTTGKLPVPWPT 60

QY     51  LVTLTSGVQCFSRYPDHMKQHDFFKSAPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db     51  LVTLTSGVQCFSRYPDHMKQHDFFKSAPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY     121 VNRIELKGIDFKEDGNILGHKLEYNYNHNHVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db     121 VNRIELKGIDFKEDGNILGHKLEYNYNHNHVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY     131 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMVLLGFVTAAGITLGMDELYK 239
Db     131 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMVLLGFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:09:27
Job time : 34.7778 secs
```



A;Contents: annotation; X-ray crystallography, 1.9 angstroms  
C;Comment: This protein is excited by the photoprotein aequorin (see PIR:AQJFNV) emitting  
C;Comment: The chromophore of this protein is formed by modification of Ser-dehydro-Tyr-  
C;Genetics:

A;Gene: GPP  
A;Introns: 59/3; 167/3  
C;Superfamily: green-fluorescent protein  
C;Keywords: Chromoprotein; luminescence  
F;65-67/Cross-link: 5-imidazolinone (Ser-Gly) #status experimental  
F;66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match 97.0%; Score 1235; DB 1; Length 238;  
Best Local Similarity 96.6%; Pred. No. 1.6e-96;  
Matches 230; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVELDGVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 61  
DB 1 MSKGEELFTGVVPIVLVELDGVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 60  
QY 52 VTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNKYKTRAEVKFEGDTLV 121  
DB 51 VTTFSGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIIFKDDGNKYKTRAEVKFEGDTLV 120  
QY 122 NRIELAGIDFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFIRHNIEDGSVOLAD 181  
DB 121 NRIELAGIDFKEDGNILGHKMEYNHSHVYIMADKQNGIKVNFIRHNIEDGSVOLAD 180  
QY 192 HYQONTPIGDGPVLLPDNHVLSQTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239  
DB 131 HYQONTPIGDGPVLLPDNHVLSQTQSALSKDPNEKRDHMLLEFVTAAGITGMDELYK 238

RESULT 2  
H72228  
hypothetical protein TM1624 - Thermotoga maritima (strain MSB8)  
C;Species: Thermotoga maritima  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
A;Accession: H72228  
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.

Nature 399, 323-329, 1999  
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A;Reference number: A72200; MUID:99287316; PMID:10360571  
A;Accession: H72228

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-785 <ARN>  
A;Cross-references: GB:AE001806; GB:AE000512; NID:g4982196; PIDN:AAD36691.1; PID:g498219  
A;Experimental source: strain MSB8  
C;Genetics:  
A;Gene: TM1524

Query Match 8.2%; Score 105; DB 2; Length 785;  
Best Local Similarity 19.7%; Pred. No. 0.84;  
Matches 46; Conservative 32; Mismatches 71; Indels 84; Gaps 7;

QY 3 SKGEELFTGVVPIVLVELDGVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTLV 62  
DB 15 NEGRFSEGTVPGVQAD-----LVRKGLLPHYVGM- 46  
QY 53 TLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNKYKTRAEVKFEGDTLVN 122  
DB 47 -----NEDLFKEIDREWIYVERFEKDEKGERVDLVEFGVDTL 88  
QY 123 RIELKIDFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFIRHNIEDGSVOLADH 182  
DB 39 DVYLTNGVYL--GSTEDMFIEYRFDVTNL-----KEKNHLKVYIK-----SPIRVPKT 134  
QY 193 YQONTPIGDGPVLLPDNHVLSQTQSALSKDPNEKRDHMLVGFVTAAGITLGM 235  
DB 135 LEQNYGLVGGP-----EDP-----INGYIRKAQYSYGD 163

RESULT 3

H64102

leucine-tRNA ligase (EC 6.1.1.4) - Haemophilus influenzae (strain Rd KW20)

N;Alternate names: leucyl-tRNA synthetase  
C;Species: Haemophilus influenzae

C;Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 03-Jun-2002  
C;Accession: H64102

R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.

Science 269, 496-512, 1995  
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, C.

A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A;Reference number: A64000; MUID:95350630; PMID:7542800

A;Accession: H64102

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-861 <TIGR>

A;Cross-references: GB:U32774; GB:L42023; NID:g1573942; PIDN:AAC22581.1; PID:g1573943; T3

C;Genetics:

A;Gene: leuS

C;Superfamily: leucine-tRNA ligase

C;Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 7.6%; Score 96.5; DB 2; Length 861;  
Best Local Similarity 24.6%; Pred. No. 4.9;  
Matches 47; Conservative 26; Mismatches 59; Indels 59; Gaps 11;

QY 50 TTGKLPVWPPTLVTLSSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNKYKTRAEVKFEGDTLV 103  
DB 314 TGDKLPI-WVANFVLMHYGTGAVMAVPAH-DQRDF-----EFAQKSLPIKQVIAPLA 364  
QY 104 DGNKYKTRAEVKFEGDTLVNRIELKIDFKEDGNILGHKLEYNHSHVYIMADK-QKNGI 162  
DB 365 DEEIDLTQKAFVEHGLKLVNSDFDGKNF--DGAENG-----IADKLEKLG 408  
QY 163 ---KVNFKIRH-----NIEDGSVOLADHYQQNTPIGDGPVLLPDNHYL- 202  
DB 409 GKQVNYRLRDGWVSQRVYWGAPIPMLTLENGDVPA-----PMEDLPILPEDVMD 461

RESULT 4

E82590

leucyl-tRNA synthetase XF2176 [imported] - Xylella fastidiosa (strain 9a5c)

C;Species: Xylella fastidiosa

C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
C;Accession: E82590

R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A;Reference number: A82515; MUID:20365717; PMID:10910347

A;Note: for a complete list of authors see reference number A59328 below

A;Accession: E82590

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-887 <SIM>

A;Cross-references: GB:AE004031; GB:AE003849; NID:g9107309; PIDN:AAF84975.1; GSPDB:GN0012

A;Experimental source: strain 9a5c

R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H.

as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.

A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Wenck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki

A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A;Reference number: A59328  
A;Content: annotation  
C;Genetics:  
A;Gene: XP2176  
C;Superfamily: leucine-tRNA ligase

Query Match 7.4%; Score 94.5; DB 2; Length 887;  
Best Local Similarity 22.7%; Pred. No. 7.5;  
Matches 45; Conservative 29; Mismatches 67; Indels 57; Gaps 10;

QY 50 TTGKLPVMPPTLTLVLSYGVCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDGNY-- 107  
Db TNEQLPV-WVANFVLMAYGTGAVMAVPGHQDQDEF--ANKYGLPIRQVIALKEPKNQDB 385

QY 108 -----KTRAEVKFEGDTLVNRIELKGDIDFKEDGNILGHKLEYNNSHNVYI 153  
Db STWEPDVRDWDYADKTR--EFE---LINSAPDGLDYQDAFEVLAERFE----- 429

QY 154 MADQKNG-IKNFKIRHNIEDGSVQLADHYQONTPI-----GDGPVLLPDN 199  
Db ---RQGRQRRVNYRLR----DWGVSQRQYWGCPFIVVYCTCGAVVPEDQLPVILPEN 482

QY 200 -HYLSQTQSALSADPNR 216  
Db VAPSGTGSPIKTPDWRK 500

RESULT 5  
leucyl-tRNA synthetase [imported] - Salmonella enterica subsp. enterica serovar Typhi (S  
C;Species: Salmonella enterica subsp. enterica serovar Typhi  
A;Note: this species has also been called Salmonella typhi  
C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
R;Accession: AC0582  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A;Reference number: AB0502; MUID:21534947; PMID:11677608  
A;Accession: AC0582  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-860 <PAR>  
A;Cross-references: GB:AL513382; PIDN:CAD05125.1; PID:g16501899; GSPDB:GN00176  
C;Genetics:  
A;Gene: STY0699  
C;Superfamily: leucine-tRNA ligase

Query Match 7.3%; Score 92.5; DB 2; Length 860;  
Best Local Similarity 23.9%; Pred. No. 11;  
Matches 43; Conservative 21; Mismatches 77; Indels 39; Gaps 7;

QY 50 TTGKLPVMPPTLTLVLSYGVCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDGNYKT 109  
Db TGEIIPV-WVANFVLMAYGTGAVMAVPGH-DQKDD-YEFASKYGLTIKPVILLADGSEPL 370

QY 110 RAEVKFEGDTLVNRIELKGDIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIR 169  
Db SEQALTEKGVLFNSGFDFGLAEFAAFNAIADKL-----AEKVGGERKVNRYLR 418

QY 170 H-----NIEDGSVQLADHYQONTPIGDGPVLLPDNHYL-STQSALSADPN 212  
Db DWGVSQRQYWGAPIMVTLDEGT-----LPTPEDQLPVILPDVDMVDGITSPIKADP 471

RESULT 6  
DNA topoisomerase IV subunit B BH2140 [imported] - Bacillus halodurans (strain C-125)  
C;Species: Bacillus halodurans

C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
R;Accession: D83917  
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hiran  
Nucleic Acids Res. 28, 4317-4331, 2000  
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A;Reference number: A83650; MUID:20512582; PMID:11058132  
A;Accession: D83917  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-655 <STO>  
A;Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BA05859.1; GSPDB:GN001  
C;Genetics:  
A;Experimental source: strain C-125  
C;Gene: BH2140  
C;Superfamily: DNA topoisomerase (Atp-hydrolyzing) chain B

Query Match 6.9%; Score 87.5; DB 2; Length 655;  
Best Local Similarity 21.9%; Pred. No. 20;  
Matches 53; Conservative 37; Mismatches 95; Indels 57; Gaps 9;

QY 22 DVNGHK---FSVSGEGEGDAT---YGLTLTKFI-----CTTGKLPVPWP 59  
Db 63 NVTIHKQDSVVRDEGRGMPGTGHHKLGKPTPEVILTVLHAGGKFGGQGYATSGLHGVA 122

QY 60 TLVTLISYGVCFSRYPDHMKQHDFFKSAPEGYVOER-----TIFKDDG----- 105  
Db 123 SVVNALSEWLIVEIKRDGWVYEQRPENGCKPSTTLEKKGKTRGTGTHFKPDPVTFSTT 182

QY 106 --NYKTRAEVKFEGDTLVNRIELKGDIDFKEDGNILGHKLEYNNSHNVIMADK----- 157  
Db 183 NFNVETLSERLEAAFLKGLKIELVLDLDDTKEVPH-YEDGKAFVEYLNEKETHLPV 241

QY 158 -----QKNGIKVNEKIRHNIEDGSVQLADHYQONTPIGDGPVLLPDNHYLSTQSALSADPN 212  
Db 242 VFNESNGIEIEFAPQFN--DGYENVLSFVNVNRTKDG-----GTHELGAKTAMTRAV 294

QY 213 NE 214  
Db 295 NE 296

RESULT 7  
D71614  
hypothetical protein PF0460c - malaria parasite (Plasmodium falciparum)  
C;Species: Plasmodium falciparum  
C;Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
C;Accession: D71614  
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
Perte, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;  
Science 282, 1126-1132, 1998  
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A;Reference number: A71600; MUID:99021743; PMID:9804551  
A;Accession: D71614  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-2573 <GAR>  
A;Cross-references: GB:AE001396; GB:AE001362; NID:g3845188; PIDN:AAC71881.1; PID:g384519  
A;Experimental source: clone 3D7  
C;Genetics:  
A;Gene: PF0460c

Query Match 6.9%; Score 87.5; DB 2; Length 2573;  
Best Local Similarity 26.2%; Pred. No. 1.2e+02;  
Matches 34; Conservative 30; Mismatches 53; Indels 13; Gaps 5;

QY 94 VOERTIFFKDD--DGNKYKTRAEVKFEGDTLVNRIELKGDIDFKEDGNILGHKLEYN--YNH 149  
Db 126 LKKTILCKDIKSGNDPMDIEISLFKDDMVDDKELK--DFEKSSSLKIKNKEVNFYNNK 183

QY 150 NYTIMADKQNGIKVNFKIRHNIEDGSVQLADHYQONTPIGDGPVLLPDNHYLSTQSALS 209  
Db 184 NLHAIKENKKDKKKNKHNNNDNNM----IYYKNI---DKTHYILDNNVVHILNDIN 236

QY 210 KDPNEKRDM 219  
Db 237 TYLKRERYM 246

RESULT 8  
E70390  
N:Contains: iron-sulfur cofactor synthesis protein nifs - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 07-Dec-1999  
C:Accession: E70390  
C:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666; PMID:9537320  
A:Accession: E70390  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-370 <AQF>  
A:Cross-references: GB:AE000720; NID:g2983529; PIDN:AAC07111.1; PID:g2983536; GB:AE00065  
A:Experimental source: strain VF5  
C:Genetics:  
A:Gene: nifs1  
C:Superfamily: nitrogen fixation protein nifs  
C:Keywords: phosphoprotein; pyridoxal phosphate; sulfurtransferase  
F:195/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted  
F:318/Active site: Cys (cysteine persulfide intermediate) #status predicted

Query Match 6.8%; Score 86.5; DB 2; Length 370;  
Best Local Similarity 25.4%; Pred. No. 11;  
Matches 49; Conservative 30; Mismatches 91; Indels 23; Gaps 8;

QY 4 KGBELFTGVV----PILVELD----GDVNGHKF-SVSGEG----EGDATYKGLTLFICT 50  
Db 164 KGVPLLTDAVQAGKPIELKNISYATFSQGHKFAIKSGSFLYISDEANVEPLIVGGQE 223

QY 51 TGKLP-----VPMTLVTLLSYGVQCFSPYDPMKQ-HDFFKSAMPEGVQERTIFPKDD 104  
Db 224 NKRSGTENNVGLISLAKALEIIVNSFQYQEQKKLRDLFENLLEA-LPDAQIVGKOA 282

QY 105 GNYKTRAEV---KFGDGLNRELKGDPKEDGNILGHKLYNNSHNHYIMADKQNG 161  
Db 283 ERSPSISVIMPFFGAIEIVNKLSEKGIYCSGACLSGEYEFNKHMLXMGFSQERALRM 342

QY 162 IKVNFKIRHNIED 174  
Db 343 VRFSFGLNKEEE 355

RESULT 9  
JDVLJHH  
DNA-directed DNA polymerase (EC 2.7.7.7) - heron hepatitis virus  
C:Species: heron hepatitis virus, HHV9  
A:Note: host Ardea cinerea (gray heron)  
C:Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 11-Jun-1999  
C:Accession: A30082  
R:Sprengel, R.; Kaleta, E.F.; Will, H.  
J. Virol. 62, 3832-3839, 1988  
A:Title: Isolation and characterization of a hepatitis B virus endemic in herons.  
A:Reference number: A93037; MUID:98333160; PMID:3418788  
A:Accession: A30082  
A:Molecule type: DNA  
A:Residues: 1-788 <SPR>  
A:Cross-references: GB:M22056; NID:g325452; PIDN:AAA45738.1; PID:g325454  
C:Superfamily: hepatitis virus DNA-directed DNA polymerase  
C:Keywords: DNA biosynthesis; nucleotidyltransferase

Query Match 6.8%; Score 86.5; DB 1; Length 788;  
Best Local Similarity 19.7%; Pred. No. 30;  
Matches 40; Conservative 32; Mismatches 66; Indels 65; Gaps 10;

QY 58 WPTLVTLLSYGVQCFSPYDPMKQH-----DFFKSAMPEGVQERT-----IFFKDDGNKY 108  
Db 139 WPKSISYLPVHSVGVPKYPFQQHESLVNDYLNKLFAGILYKVSXHLVTFK--GPVF 196

QY 109 T-----RAEVKPEGDTLVNRIELKGDIDFKEDGNILGHKLYNNSHNHYIM 154  
Db 197 TWEQKHLVPQOHGAYSSKINDRQESRRRIITATSSRKNDSSRI-----FGAHN----- 245

QY 155 ADKQKNGIKVNFKIRHNIEDSGVQLADHYQ-----QNTPIGDGPVLL--PDNHYL 202  
Db 246 -----NGRKISY---HSTRDGHSLRSTSDPTSRGALAGDSTPIGPGSTAHPSTHHV 297

QY 203 STQ-----SALSKDPNEKR 216  
Db 298 DRRRQKQGVQLAISREPSETR 320

RESULT 10  
JC4078  
Protective surface antigen D-15 precursor - Haemophilus influenzae (type b)  
C:Species: Haemophilus influenzae  
A:Variety: Type b  
C:Date: 30-Jun-1995 #sequence\_revision 14-Jul-1995 #text\_change 29-Sep-1999  
C:Accession: JC4078  
R:Flack, F.S.; Loosmore, S.; Chong, P.; Thomas, W.R.  
Gene 156, 97-99, 1995  
A:Title: The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus infl.  
A:Reference number: JC4078; MUID:95255676; PMID:7737523  
A:Accession: JC4078  
A:Molecule type: DNA  
A:Residues: 1-797 <FLA>  
A:Cross-references: GB:U13961; NID:g537447; PIDN:AAA85645.1; PID:g537448  
A:Experimental source: type b  
C:Superfamily: protective surface antigen D-15  
C:Keywords: surface antigen  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-797/Product: protective surface antigen D-15 #status predicted <MAT>

Query Match 6.8%; Score 86.5; DB 2; Length 797;  
Best Local Similarity 21.9%; Pred. No. 31;  
Matches 48; Conservative 30; Mismatches 78; Indels 63; Gaps 11;

QY 65 LSYGVQCFSPYDPMKQHDF-----FKSAMPEGVQEE-----RTI 99  
Db 427 IGVGTESGISYQASVKQDNFLGTGAASVSIAGTKNDYGTGTVNLGYTEPYFTKDGVSJLGSNV 486

QY 100 FFKDDGNKYTRAEVKPEGDTLVNRIELKGDIDFKEDGNI---LGH-----KLEYNNS 148  
Db 487 FFENYDNSKSDTSSNRYKRTYGSNVTL-GFPVNNNSYVVGLOHTYKNKISNFALEYN--- 542

QY 149 HNVYIMADKQK-NGIKVNFKIRHNIEDSGVQLADHYQ-----NTPIGDGPVLL 196  
Db 543 RNLIYQSMKFGNGIKTN-----DFDFSFGVNNLSLRNGYFPTKGVKASLG-GRVTI 593

QY 197 P--DNHYLSTQSAKSKDPNEKRDHVMVLLGLFVTAAGITLG 233  
Db 594 PGSDNKKYKLSADVOGFYPLDRDLHLMVWSAKASAGYANG 632

RESULT 11  
F64102  
Protective surface antigen D-15 - Haemophilus influenzae (strain Rd KW20)  
C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 18-Sep-1998  
C:Accession: F64102  
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,  
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630; PMID:7542800  
A:Accession: F64102

Db 272 -EIVNENAFVLSNDWGSNNIRLAMYIGENGYATNPEVK-----DLVVEGIELA----- 319

Db 272 -EIVNENAFVNSNDWGSNMIRLAMYIGENCYATNPVK---DLVYEGIELA----- 319

A;Title: Molecular cloning and nucleotide sequence of the gene for *Escherichia coli* leucine aminopeptidase  
A;Reference number: A30290; MUID:88096562; PMID:3320963

A;Accession: A30290  
A;Molecule type: DNA  
A;Residues: 1-66, 'H', 68-195, 'N', 197-261, 'R', 263-860 <HAE>  
A;Cross-references: EML:X06331; NID:G41915; PIDN:CAA29642.1; PID:G41916  
A;Note: part of this sequence, including the amino end of the mature protein, was confirmed by mass spectrometry.  
C;Genetics:  
A;Gene: leuS  
A;Map position: 15 min  
C;Function:  
A;Description: EC 6.1.1.4 [validated, MUID:88096562]  
A;Pathway: protein biosynthesis  
C;Superfamily: leucine-tRNA ligase  
C;Keywords: aminoacyl-tRNA synthetase; ATP; ligase; protein biosynthesis  
F;49-52/Region: ATP binding #status predicted

Query Match 6.7%; Score 85.5; DB 1; Length 860;  
Best Local Similarity 22.8%; Pred. No. 41;  
Matches 41; Conservative 24; Mismatches 76; Indels 39; Gaps 8;

QY 50 TTGKLPVPWPTLVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKT 109  
DB 314 TGEIIPV-WAANFVLMVEYGTGAVNVPFGH-DQRD-YEFASKYGLNFKPVILADGSEPD 370  
QY 110 RAEVKFEGDGLVNRLEIKGIDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIR 169  
DB 371 SQQALTEKGVLFNSGEFGLD-----HEAFAAIADKLTAMGVGER---KVNRYLR 418  
QY 170 H-----NIEGVSQVLADHYQQNTPTGDPVLLPDNHYL-STQSALSKDP 212  
DB 419 DWGVSQRQYWGAPIPMTVLEDGTV-----MPTDDQLPVILPEDVVMGDGITSPIKADP 471

RESULT 15  
H90713  
leucine tRNA synthetase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 050952)  
C;Species: Escherichia coli  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C;Accession: H90713  
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome analysis of the associated hemorrhagic colitis outbreaks in Japan.  
A;Reference number: A99629; MUID:21156231; PMID:11258796  
A;Accession: H90713  
A;Status: Preliminary  
A;Molecule type: DNA  
A;Residues: 1-860 <HAY>  
A;Cross-references: GB:BA000007; PIDN:BAE34103.1; PID:gl3360138; GSPDB:GN00154  
A;Experimental source: strain O157:H7, substrain RIMD 050952  
C;Genetics:  
A;Gene: ECs0680  
C;Superfamily: leucine-tRNA ligase

Query Match 6.7%; Score 85.5; DB 2; Length 860;  
Best Local Similarity 22.8%; Pred. No. 41;  
Matches 41; Conservative 24; Mismatches 76; Indels 39; Gaps 8;

QY 50 TTGKLPVPWPTLVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKT 109  
DB 314 TGEIIPV-WAANFVLMVEYGTGAVNVPFGH-DQRD-YEFASKYGLNFKPVILADGSEPD 370  
QY 110 RAEVKFEGDGLVNRLEIKGIDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIR 169  
DB 371 SQQALTEKGVLFNSGEFGLD-----HEAFAAIADKLTAMGVGER---KVNRYLR 418  
QY 170 H-----NIEGVSQVLADHYQQNTPTGDPVLLPDNHYL-STQSALSKDP 212  
DB 419 DWGVSQRQYWGAPIPMTVLEDGTV-----MPTDDQLPVILPEDVVMGDGITSPIKADP 471

Search completed: June 21, 2004, 16:01:57  
Job time : 11.3333 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:42:24 ; Search time 6.4444 Seconds  
(without alignments)  
1931.085 Million cell updates/sec

Title: US-09-887-784-64L

Perfect score: 1273

Sequence: 1 MVSKGEELFTGVVPLVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1241	97.5	238	1	GFP_AEQVI
2	96.5	7.6	861	1	SYL_HAEIN
3	94.5	7.4	879	1	SYL_XYLFA
4	92.5	7.3	860	1	SYL_SALTI
5	92.5	7.3	860	1	SYL_SALTY
6	91.5	7.2	879	1	SYL_XYLFT
7	89.5	7.0	859	1	SYL_SHEON
8	88.5	7.0	538	1	GRBE_RAT
9	86.5	6.8	533	1	CP51_CANGA
10	86.5	6.8	788	1	DPOL_HPBHE
11	86.5	6.8	795	1	D152_HAEIN
12	86.5	6.8	797	1	D151_HAEIN
13	86.5	6.8	941	1	GUN_BACS6
14	86	6.8	357	1	TRMF_CAMJE
15	85.5	6.7	793	1	D153_HAEIN
16	85.5	6.7	860	1	SYL_ECO57
17	85.5	6.7	860	1	SYL_ECOL6
18	85.5	6.7	860	1	SYL_ECOLI
19	85.5	6.7	886	1	ITH3_MSAU
20	85	6.7	461	1	PSBC_CVAPA
21	84.5	6.6	501	1	AMPA_WIGBR
22	84.5	6.6	613	1	PEPF_MYCPU
23	84.5	6.6	1603	1	V174_CAEEL
24	84	6.6	439	1	SV62_DISOM
25	84	6.6	504	1	YC03_KLEPN
26	84	6.6	752	1	NECI_RAT
27	83.5	6.6	860	1	SYL_YERPE
28	83.5	6.6	1164	1	BAG_STRAG
29	82	6.4	336	1	YD4E_METJA
30	82	6.4	353	1	HIS7_BUCAI
31	82	6.4	366	1	SE17_HUMAN
32	82	6.4	774	1	AMY2_SCHPO
33	81.5	6.4	589	1	SYD_HABDU
					P42212 aequorea vi
					P43827 haemophilus
					Q9pbg8 xylella fas
					Q8z8h5 salmonella
					Q8zqz6 salmonella
					Q87c65 xylella fas
					Q8ehp4 shewanella
					Q88900 rattus norv
					P50859 candida gla
					P13846 heron hepat
					P44935 haemophilus
					P46024 haemophilus
					P19424 bacillus sp
					Q9pp92 campylobact
					O32629 haemophilus
					Q8xbn8 escherichia
					Q8fiy9 escherichia
					P07813 escherichia
					P97280 mesocricetu
					P48104 cyanophora
					Q8d295 wiggleswort
					Q98gp0 mycoplasma
					P18947 caenorhabdi
					P24506 discopoge o
					Q48449 klebsiella
					P28840 rattus norv
					Q8zdf8 yersinia pe
					P27951 streptococ
					Q58743 methanococ
					P57203 buchnera sp
					Q8wt66 homo sapien
					O42918 schizosacch
					Q7vnf0 haemophilus

RESULT 1  
GFP\_AEQVI STANDARD; PRT; 238 AA.  
ID AC P42212; Q17104; Q27903;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Green fluorescent protein.  
GN GFP.  
OS Aequorea victoria (Jellyfish).  
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;  
OC Aequoreidae; Aequorea.  
OX NCBI\_TaxID=6100;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=92175527; PubMed=1347277;  
RA Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,  
RA Cormier M.J.;  
RT "Primary structure of the Aequorea victoria green-fluorescent  
RT protein.";  
RL Gene 111:229-233(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94185810; PubMed=8137953;  
RA Inouye S., Tsuji F.I.;  
RT "Aequorea green fluorescent protein. Expression of the gene and  
RT fluorescence characteristics of the recombinant protein.";  
RN [3]  
RX FEBS Lett. 341:277-280(1994).  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97299832; PubMed=9154981;  
RA Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;  
RT "Enhanced expression in tobacco of the gene encoding green fluorescent  
RT protein by modification of its codon usage.";  
RL Plant Mol. Biol. 33:989-999(1997).  
RN [4]  
RP CHROMOPHORE.  
RX MEDLINE=93192221; PubMed=8448132;  
RA Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;  
RT "Chemical structure of the hexapeptide chromophore of the Aequorea  
RT green-fluorescent protein.";  
RL Biochemistry 32:1212-1218(1993).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
RX MEDLINE=96355665; PubMed=8703075;  
RA Ormoe M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,  
RA Remington S.J.;  
RT "Crystal structure of the Aequorea victoria green fluorescent  
RT protein.";  
RL Science 273:1392-1395(1996).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
RX MEDLINE=98294543; PubMed=9631087;  
RA Yang F., Moss L.G., Phillips G.N. Jr.;  
RT "The molecular structure of green fluorescent protein.";  
RL Nat. Biotechnol. 14:1246-1251(1996).

#### ALIGNMENTS

34 81.5 6.4 2222 1 DPOE YEAST  
35 81 6.4 682 1 PRC\_ECOLI  
36 81 6.4 689 1 AC2L HUMAN  
37 81 6.4 874 1 SLAP\_BACLI  
38 81 6.4 1224 1 COPA\_HUMAN  
39 80.5 6.3 593 1 TRMB\_HELPY  
40 80.5 6.3 658 1 ADAS\_HUMAN  
41 80 6.3 737 1 OPT1\_DROME  
42 80 6.3 836 1 DPOL\_HPBDU  
43 80 6.3 887 1 ITH3\_RAT  
44 79.5 6.2 312 1 TRXB\_CHLMU  
45 79.5 6.2 365 1 IAZ5\_HUMAN

P21951 saccharomyc  
P23865 escherichia  
Q9nub1 homo sapien  
P49052 bacillus li  
P53621 homo sapien  
O25443 helicobacte  
O00116 homo sapien  
P91679 drosophila  
P03162 duck hepati  
O63416 rattus norv  
Q9pkt7 chlamydia m  
P18462 homo sapien

[7]  
 X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.  
 MEDLINE=98455509; PubMed=9782051;  
 Wachter R.M., Eislinger M.A., Kallio K., Hanson G.T., Remington S.J.;  
 "Structural basis of spectral shifts in the yellow-emission variants  
 of green fluorescent protein.";  
 Structure 6:1267-1277(1998).  
 [8]  
 X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 MEDLINE=99238303; PubMed=10220315;  
 Eislinger M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J.;  
 "Structural and spectral response of green fluorescent protein  
 variants to changes in pH.";  
 Biochemistry 38:5296-5301(1999).  
 -!- FUNCTION: Energy-transfer acceptor. Its role is to transduce the  
 blue chemiluminescence of the protein aequorin into green  
 fluorescent light by energy transfer. Fluoresces in vivo upon  
 receiving energy from the Ca(2+)-activated photoprotein aequorin.  
 Absorbs light maximally at 395 nm and exhibits a smaller  
 absorbance peak at 470 nm. The fluorescence emission spectrum  
 peaks at 509 nm with a shoulder at 540 nm.  
 -!- SUBUNIT: Monomer.  
 -!- TISSUE SPECIFICITY: Photocytes.  
 -!- PTM: Contains a covalently attached chromophore, which is composed  
 of modified amino acid residues. The chromophore is formed upon  
 cyclization of the residues Ser-dehydroTyr-Gly.  
 -!- BIOTECHNOLOGY: Has become a useful and ubiquitous tool for making  
 chimeric proteins of GFP linked to other proteins where it  
 functions as a fluorescent protein tag. GFP tolerates N- and C-  
 terminal fusion to a broad variety of proteins. It has been  
 expressed in bacteria, yeast, slime mold, plants, Drosophila,  
 zebrafish, and in mammalian cells. As a noninvasive fluorescent  
 marker in living cells, it allows for a wide range of applications  
 where it may function as a cell lineage tracer, reporter of gene  
 expression, or as a measure of protein-protein interactions.  
 -!- DATABASE: NAMB-Protein Spotlight;  
 NOTE=Issue 11 of June 2001;  
 WWW="http://www.expasy.org/spotlight/articles/sptlt011.html".  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; M62654; AAA27722.1; -;  
 EMBL; M62653; AAA27721.1; -;  
 EMBL; L29345; AAA58246.1; -;  
 EMBL; X96418; CAA65278.1; -;  
 FIR; J30692; JQ1514.  
 DR PDB; 1B9C; 17-NOV-00.  
 DR PDB; 1BFP; 07-JUL-97.  
 DR PDB; 1CAF; 14-JUN-00.  
 DR PDB; 1EWA; 08-NOV-96.  
 DR PDB; 1EMB; 16-JUN-97.  
 DR PDB; 1EMC; 20-AUG-97.  
 DR PDB; 1EME; 20-AUG-97.  
 DR PDB; 1EMF; 20-AUG-97.  
 DR PDB; 1EMG; 12-MAY-99.  
 DR PDB; 1EMK; 20-AUG-97.  
 DR PDB; 1EML; 20-AUG-97.  
 DR PDB; 1EMW; 20-AUG-97.  
 DR PDB; 1F09; 17-NOV-00.  
 DR PDB; 1F0B; 17-NOV-00.  
 DR PDB; 1GEL; 11-JAN-97.  
 DR PDB; 1H03; 15-JAN-02.  
 DR PDB; 1HUY; 04-JUL-01.  
 DR PDB; 1JBY; 07-JAN-03.  
 DR PDB; 1JBZ; 07-JAN-03.  
 DR PDB; 1KP5; 28-AUG-02.  
 DR PDB; 1KVP; 10-APR-02.

DR PDB; 1KVR; 10-APR-02.  
 DR PDB; 1KYS; 10-APR-02.  
 DR PDB; 1YFP; 28-OCT-98.  
 DR PDB; 2EMD; 20-AUG-97.  
 DR PDB; 2EMN; 20-AUG-97.  
 DR PDB; 2EMO; 20-AUG-97.  
 DR InterPro: IPR009017; GFP like.  
 DR InterPro: IPR000786; Green\_fl\_protein.  
 DR Pfam: PF01353; GFP; 1.  
 DR PRINTS; PD01229; GFLUORESCENT.  
 DR ProDom; PD013756; Green\_fl\_protein; 1.  
 KW Luminescence; 3D-structure.  
 FT CROSSLINK 65 67 5-imidazolinone (Ser-Gly).  
 FT MOD\_RES 66 66 2,3-DIDEHYDROTYROSINE.  
 FT VARIANT 100 100 F -> Y.  
 FT VARIANT 108 108 T -> S.  
 FT VARIANT 141 141 L -> M.  
 FT VARIANT 219 219 V -> I.  
 FT CONFLICT 2 2 S -> G (IN REF. 3).  
 FT CONFLICT 25 25 H -> Q (IN REF. 2).  
 FT CONFLICT 80 80 Q -> R (IN REF. 3).  
 FT CONFLICT 157 157 Q -> P (IN REF. 2).  
 FT CONFLICT 172 172 E -> K (IN REF. 2).  
 FT HELIX 4 8  
 FT STRAND 12 22  
 FT TURN 23 24  
 FT STRAND 25 36  
 FT TURN 37 40  
 FT STRAND 41 48  
 FT TURN 49 50  
 FT HELIX 57 60  
 FT TURN 61 63  
 FT HELIX 69 71  
 FT STRAND 73 73  
 FT HELIX 76 81  
 FT HELIX 83 86  
 FT TURN 87 90  
 FT STRAND 92 100  
 FT TURN 101 102  
 FT STRAND 105 115  
 FT TURN 116 117  
 FT STRAND 118 128  
 FT TURN 132 133  
 FT TURN 135 139  
 FT STRAND 141 141  
 FT STRAND 148 155  
 FT TURN 156 159  
 FT STRAND 160 171  
 FT TURN 172 173  
 FT STRAND 176 187  
 FT STRAND 199 208  
 FT TURN 211 212  
 FT STRAND 217 227  
 SQ SEQUENCE 238 AA; 26886 MW; EA5A6F21FBFB6E05 CRC64;  
 Query Match 97.5%; Score 1241; DB 1; Length 238;  
 Best Local Similarity 97.9%; Pred.No.3.3e-95;  
 Matches 233; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVWPPTL 61  
 :|||||  
 Db 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVWPPTL 60  
 QY 62 VTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 121  
 :|||  
 Db 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 120  
 QY 122 NRIELKGIDFKEDGNILGHLEINYNSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181  
 :|||||  
 Db 121 NRIELKGIDFKEDGNILGHLEINYNSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180  
 QY 182 HYQQNTPIGDGPVLLPDNHVLTQSALSKDPNEKDHMVLLGFVTAAGTITLGMDELYK 239  
 :|||||

```
Db 181 HYQONTPIGDPVLLPDNHYLSQTSALSQKDPNEKRDHMLVLLBFVTAAGITHGMDELYK 238
RESULT 2
SYL_HAEIN
ID SYL_HAEIN STANDARD; PRT; 861 AA.
AC P43827;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--trNA ligase) (LeuRS).
GN LEUS OR HI0921.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd."
RL Science 269:496-512(1995).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; U32774; AAC22581.1; -
DR PIR; H64102; H64102.
DR TIGR; HI0921; -
DR HAMAP; MF_00049; -; 1.
DR InterPro; IPR002302; Leu-trNASynt1a.
DR InterPro; IPR022300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR009008; ValRS_fliers_edit.
DR Pfam; PF00133; tRNA-synt 1; 1.
DR PRINTS; P000985; TRNASYNTHLEU.
DR TIGRfams; TIGR00396; leuS bact; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY)
SQ SEQUENCE 861 AA; 97750 MW; EB93304F6B4C8FB7 CRC64;
Query Match 7.6%; Score 96.5; DB 1; Length 861;
Best Local Similarity 24.6%; Pred. No. 2.6;
Matches 47; Conservative 26; Mismatches 59; Gaps 11;
QY 50 TTGKLPVNPPTLVTLLSYGVCFSRYPDHMKQHDFFKSAWPEGYQVERIFFKD----- 103
Db 314 TGDKLPI-WVANFVLMHYGTGVMAVPAH-DQRDF-----EFAQKYSLPKIQVIAPLA 364
```

```
QY 104 DQNYKTRAEVKEGDTLVNRIELKGIQFKEDGNIILGHKLEYNYNHNVIMADK-QKNGI 162
Db 365 DEEIDLTKQAFVEHGLKLVNSDFDGKNF--DGAFNG-----IADKLEKLGV 408
QY 163 ---KYNFKIRH-----NIEDSGSVQLADHYQQNTPIGDPVLLPDNHYL- 202
Db 409 GKRQVNYRLRDWGVSRQRYWGAPIPMLTLLENGDVVPA-----PWEDLPILPEDVWMD 461
QY 203 STQSALSQKDPN 213
Db 462 GVKSPINADPN 472
RESULT 3
SYL_XYLFA
ID SYL_XYLFA STANDARD; PRT; 879 AA.
AC Q9P68;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--trNA ligase) (LeuRS).
GN LEUS OR XF2176.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5C; PubMed=10910347;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Cartaro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.P., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincini A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hobeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Leite L.C.C.,
RA Machado M.A., Madeira A.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.C., Nunes L.R., Oliveira M.A., Paris A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa."
RL Nature 406:151-159(2000).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; AF004031; AAF84975.1; ALT_INIT.
DR HAMAP; MF_00049; -; 1.
```



[illegible]

RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sasaki P.T., Sena J.A.D.,  
RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,  
RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,  
RA Kitajima J.P.;  
RT "Comparative analyses of the complete genome sequences of *Xylella*  
RT fastidiosa" and citrus variegated chlorosis strains of *Xylella*  
RT disease and citrus variegated chlorosis strains of *Xylella*  
RL J. Bacteriol. 185:1018-1026 (2003).  
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA (Leu) = AMP +  
CC diphosphate + L-leucyl-tRNA (Leu).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AB012557; AAC29080.1; ALT\_INIT.  
CC HAMAP: MF 00049; -; 1.  
CC InterPro: IPR002302; Leu-tRNA-synt1a.  
CC InterPro: IPR002300; tRNA-synt\_1a.  
CC InterPro: IPR001412; tRNA-synt\_1.  
CC InterPro: IPR009008; ValRS\_ILERS\_edit.  
CC Pfam: PF00133; tRNA-synt 1; 1.  
CC PRINTS: PR00985; TRNASYNTHLEU.  
CC TIGRfam: TIGR00396; leus\_bact; 1.  
CC PROSITE: PS00178; AA\_TRNA\_LIGASE\_I; 1.  
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
CC Complete proteome.  
CC SITE 45 55 "HIGH" REGION.  
CC SITE 637 641 "KMSKS" REGION.  
CC BINDING 640 640 ATP (BY SIMILARITY).  
CC SEQUENCE 879 AA; 99823 MW; 4C2EE01B8FDC497E CRC64;  
Query Match 7.2%; Score 91.5; DB 1; Length 879;  
Best Local Similarity 22.7%; Pred. No. 6.8;  
Matches 45; Conservative 28; Mismatches 68; Indels 57; Gaps 10;  
QY 50 TTGKLPVPWPTLVTLLSYGVQCFSRYPDHMKQHDFFKSNMPEGYVQERTIFFKDDGNYK 107  
DB 321 TNEQLP-V-WVANFVLMAYGTGAVMAVPGHQDQDEF--ANKYGLPIRQVIALKEPKNQDE 377  
QY 108 -----KTRAEVKEGDTLVNRIELKGDGDKNGLGHKLEYNNSHNYI 153  
DB 378 STWEPDVRWDYADKTR---EFE---LINSAEFDGLDYQGAFFVLAERFE----- 421  
QY 154 MADKQKNG-IKYNFKIRHNIEDSGVQLADHYQQNTPI-----GDGPVLLPDN 199  
DB 422 ---RQGRGQRRVNYRLR----DWGVSQRQYWGCPPIVYCTCGAVPVPENQLPILPEN 474  
QY 200 -HYLSTQSALSQDPNEKR 216  
DB 475 VAFSGTGSPIKTDPEWRK 492  
RESULT 7  
SYL\_SHEON  
ID SYL\_SHEON STANDARD; PRT; 859 AA.  
AC Q8EHP4;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (leucine--tRNA ligase) (LeuRS).  
GN LEUS OR SO1174.  
OS Shewanella oneidensis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
OC Alteromonadaceae; Shewanella.  
OX NCBI\_TaxID=70863;  
RN [1]

```
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealeson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE015561; AAN54244.1; -.
DR HAMAP; MF 00049; -.
DR TIGR; SO1174; -.
DR InterPro; IPR002302; Leu-tRNA-synt1a.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR009008; ValRS_1leRS_edit.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRFAMs; TIGR00396; leuS_bact; 1.
DR PROSITE; PS00178; AA TRNA_LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 618 622 "KMSKS" REGION.
FT BINDING 621 621 ATP (BY SIMILARITY).
SQ SEQUENCE 859 AA; 96827 MW; C78D6209DFB6CA17 CRC64;

Query Match 7.0%; Score 89.5; DB 1; Length 859;
Best Local Similarity 24.4%; Pred. No. 9.7;
Matches 51; Conservative 29; Mismatches 95; Indels 37; Gaps 9;

QY 18 ELDGDVNGHKFSVSGEGE-----GDATYVKLTGKLTGKLPVWPVPTLVLLSYGVQ 70
DB 277 ELAAFDICEKNSTSAELATMEKRGVAT-GLYAIHPI--TGKQVPIAANFVLMNYGTG 333
QY 71 CFSRYPDH-MKQHDFFKSPAMPEGVQERTIFFKDDGNKYKTRAEVKFEGDPLVNRIELKGI 129
DB 334 AVMSVFGHQDQRYEFAK----KVHLPLEAVIKPAEGDLDISEAAYTEKGLFNSGEGDGL 389
QY 130 DKEDGNILGHKLEYNNSHNVIMADKQNGIKVNFKIRH-----NIEDGSGVOLA 180
DB 390 DFDGAFNFIANKL-----VAGKGRK--QVNYRLRDMGVSQRQYWGAPIPMVTILA 437
QY 131 DHYQONTPTIGDGPVLLPDNHYL-STQSALSKD 211
DB 438 DGTVITPDPDLPVLPEDVMDGIQSPIKAD 469

RESULT 8
GRBE RAT
ID GRBE RAT STANDARD; PRT; 538 AA.
AC O88970;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE GN OS
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar;
RX MEDLINE=98421528; PubMed=9748281;
RA Kasus-Jacobi A., Perdereau D., Auzan C., Clauser E., van Obberghen E.,
RA Mauvais-Jarvis F., Girard J., Burnol A.-F.;
RT "Identification of the rat adapter Grb14 as an inhibitor of insulin
RT actions.";
RL J. Biol. Chem. 273:26026-26035(1998).
CC -1- FUNCTION: Interacts with the cytoplasmic domain of the
CC autophosphorylated insulin receptor which is then inhibited. The
CC interaction is mediated by the SH2 domain.
CC -1- SUBUNIT: Binds to the ankyrin repeat region of TNKL via its N-
CC terminus (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and
CC endosomes (By similarity).
CC -1- PTM: Phosphorylated on serine residues (By similarity).
CC -1- SIMILARITY: Contains 1 PH domain.
CC -1- SIMILARITY: Contains 1 Ras-associating domain.
CC -1- SIMILARITY: Contains 1 SH2 domain.
CC -1- SIMILARITY: Belongs to the GRB7/10/14 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF076619; AAC61478.1; -.
DR HSSP; P35235; IAVA.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000159; RA_domain.
DR InterPro; IPR000980; SH2.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00788; RA; 1.
DR Pfam; PF00017; SH2; 1.
DR PRINTS; PR00401; SH2DOMAIN.
DR ProDom; PD000093; SH2; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00314; RA; 1.
DR SMART; SM00252; SH2; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50200; RA_1.
DR PROSITE; PS50001; SH2; 1.
KW SH2 domain; Phosphorylation.
FT DOMAIN 104 190
FT DOMAIN 232 340 PH.
FT DOMAIN 437 533 SH2.
SQ SEQUENCE 538 AA; 60592 MW; CEEC9037E7868EEF CRC64;

Query Match 7.0%; Score 88.5; DB 1; Length 538;
Best Local Similarity 19.4%; Pred. No. 6.7;
Matches 41; Conservative 33; Mismatches 70; Indels 67; Gaps 9;

QY 41 GKLTGKLTGKLTGKLPVWPVPTLVLLSYGVQCFSRV--PDHMKQHDFFKSPAMPEGVQSR 97
DB 313 GPRDLKMLCAEEDQSRMCWVTAILRLKYGMQLYQNYMHPQSARSACSSQSVSPMRSVSEN 372
QY 98 TIFPKDDGNKYKTRAEVKFEGDPLVNRIELKGIIDFKEDG-----NIL 138
DB 373 SLVAMDFSGQKTRV-IDNPTEALSVAVE-EGLAWRKKGCLRLGNHGSPTAPSSAVNNA 430
QY 139 GHKLEYNNSH-----NVYIMADKQK-----GIKVPF----- 166
DB 431 LHRSQFWFHHRISRDQAQOLITROGQVDFVLRVDSQSNPRTFVLMSHSGQKIKHFQIIP 490

Growth factor receptor-bound protein 14 (GRB14 adapter protein).
```

```
QY 167 -----KIRHNIEDGS-----VOLADHYOQN 186
      :::::
Db 491 VEDDGEVFTLDDGHTKFTDLQLQVFEYQLN 521

RESULT 9
CP51_CANGA
ID CP51_CANGA STANDARD; PRT; 533 AA.
AC PS0859; Q02312;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 51 (EC 1.14.13.70) (CYPLI) (P450-LIAl) (Sterol 14-
DE alpha-demethylase) (Lanosterol 14-alpha demethylase) (P450-14DM).
GN ERG11 OR CYP51.
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2001-15;
RX MEDLINE=96161286; PubMed=8593007;
RA Geber A., Hitchcock C.A., Swartz J.E., Pullen F.S., Marsden K.E.,
RA Kwon-Chung K.J., Bennett J.E.;
RT "Deletion of the Candida glabrata ERG3 and ERG11 genes: effect on cell
RT viability, cell growth, sterol composition, and antifungal
RT susceptibility.";
RL Antimicrob. Agents Chemother. 39:2708-2717(1995).
RN [2]
RP SEQUENCE OF 60-473 FROM N.A.
RC STRAIN=ATCC 2001;
RX MEDLINE=95081364; PubMed=7989540;
RA Burgener-Kairuz P., Zuber J.P., Jaunin P., Buchman T.G., Bille J.,
RA Rossier M.;
RT "Rapid detection and identification of Candida albicans and
RT Torulopsis (Candida) glabrata in clinical specimens by
RT species-specific nested PCR amplification of a cytochrome P-450
RT lanosterol-alpha-demethylase (LIAL) gene fragment.";
RL J. Clin. Microbiol. 32:1902-1907(1994).
CC -!- FUNCTION: Catalyzes C14-demethylation of lanosterol which is
CC critical for ergosterol biosynthesis. It transforms lanosterol
CC into 4,4'-dimethyl cholesta-8,14,24-triene-3-beta-ol (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: Obtusifoliol + 3 O(2) + 3 NADPH = 4-alpha-
CC methyl-5-alpha-ergosta-8,14,24(28)-trien-3-beta-ol + formate + 3
CC NADP(+) + 3 H(2)O.
CC -!- PATHWAY: Ergosterol biosynthesis.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

109 T-----BAEVKFGDTLVNRILKGIQKEDGNILGHKLEYNYNHNVYIM 154  
197 TWEQKHLVPOOHGAYSKINDROESRRRIITATSRKNDSSRI-----FGAHN----- 245  
155 ADKQKNGIKVNFIRNIEDGVSQVLADHYQ-----QNTPIGDGPVLL-PDNHYL 202  
246 -----NGRKISY-----HSTRDGHSLRSGTSDPTSRGALAGDSTPIGPGSTAHPSTHHV 297  
203 STQ-----SALSQDNPKR 216  
298 DRRRQKQGVLOAIGREPSETR 320

## RESULT 11

D152\_HAEIN STANDARD; PRT; 795 AA.  
ID D152\_HAEIN STANDARD; PRT; 795 AA.  
AC P44935;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)  
DE (D-15-Ag) (Outer membrane protein D15).  
GN HI0917.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=RD / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
RT Rd.";  
RL Science 269:496-512(1995).  
CC -!- SUBCELLULAR LOCATION: Outer membrane.  
CC -!- SIMILARITY: Belongs to the surface antigen D15 family.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; U32773; AAC22575.1; -  
CC TIGR; HI0917; -  
DR InterPro; IPR000184; Bac\_surfAg\_D15.  
DR Pfam; PF01103; Bac\_surface\_Ag; 1.  
KW Antigen; Outer membrane; Signal.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 795 PROTECTIVE SURFACE ANTIGEN D15.  
SQ SEQUENCE 795 AA; 87478 MW; B85691FC2286ED44 CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 795;  
Best Local Similarity 21.9%; Pred. No. 16;  
Matches 48; Conservative 30; Mismatches 78; Indels 63; Gaps 11;

65 LSYGVQCFSRYPDHMKQHDF-----FKSAMPEGYVQE-----RTI 99  
427 IGYGTSGISYQASVKQDNFLGTGAAVSIAGTKNDYGTSNVLGYTFPYFTKDGVS LGNV 486  
100 FFXDGDGNKYTRAEVKFEGDTLVNRILKGIQKEDGNI-----LGH-----KLEYNYS 148

487 FFENVNDSKSDTSSNYKRTTYGSNVTL-GFPVNNNSYYVGLGHTYKNISNFALEYN--- 542  
149 HNVYIMADKQK-NGIKVNFIRNIEDGVSQVLADHYQQ-----NTPIGDGPVLL 196  
543 RNLTYQSMKFGKNGIKTN-----DFDFSGWNNNSLNRGYFPTKGVKASLG-GRVTI 593  
197 P--DNHYLSTQALSQDNPKRDKHMLVLLGFVTAAGITLG 233  
594 PGSDNKYYKLSADVQGFYFLDRDLHMLVWSAKASAGYANG 632

## RESULT 12

D151\_HAEIN STANDARD; PRT; 797 AA.  
ID D151\_HAEIN STANDARD; PRT; 797 AA.  
AC P46024;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)  
DE (D-15-Ag) (Outer membrane protein D15).  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=Serotype B;  
RX MEDLINE=95255676; PubMed=7737523;  
RA Flack F.S., Loomore S., Chong P., Thomas W.R.;  
RT "The sequencing of the 80-kDa D15 protective surface antigen of  
RT Haemophilus influenzae.";  
RL Gene 156:97-99(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=Winna / Serotype B, and Eagan / Serotype B;  
RX MEDLINE=97427952; PubMed=9284140;  
RA Loomore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,  
RA Klein M.H.;  
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae  
RT species and may represent a universal protective antigen against  
RT invasive disease.";  
RL Infect. Immun. 65:3701-3707(1997).  
CC -!- SUBCELLULAR LOCATION: Outer membrane.  
CC -!- SIMILARITY: Belongs to the surface antigen D15 family.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; U13961; AAA85645.1; -  
CC EMBL; U60832; AAB61974.1; -  
CC EMBL; U60833; AAB61976.1; -  
CC PIR; JC4078; JC4078.  
DR InterPro; IPR000184; Bac\_surfAg\_D15.  
DR Pfam; PF01103; Bac\_surface\_Ag; 1.  
KW Antigen; Outer membrane; Signal.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 797 PROTECTIVE SURFACE ANTIGEN D15.  
SQ SEQUENCE 797 AA; 87675 MW; 2F93DE53869AF1B CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 797;  
Best Local Similarity 21.9%; Pred. No. 16;  
Matches 48; Conservative 30; Mismatches 78; Indels 63; Gaps 11;

65 LSYGVQCFSRYPDHMKQHDF-----FKSAMPEGYVQE-----RTI 99  
427 IGYGTSGISYQASVKQDNFLGTGAAVSIAGTKNDYGTSNVLGYTFPYFTKDGVS LGNV 486







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:45:30 ; Search time 30.6667 Seconds  
(without alignments)  
2458.984 Million cell updates/sec

Title: US-09-887-784-64L  
Perfect score: 1273  
Sequence: 1 MVSKEELFTGVVILVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1237	97.2	238	Q8GHE2	Q8ghe2 azotobacter
2	1234	96.9	238	Q93125	Q93125 aequorea vi
3	1232	96.8	238	Q8GHE4	Q8ghe4 azomonas ag
4	1231	96.7	238	Q8GHE3	Q8ghe3 azotobacter
5	1199	94.2	238	Q17105	Q17105 aequorea vi
6	1184	93.0	238	Q17106	Q17106 aequorea vi
7	1079	84.8	238	Q8WTC6	Q8wtc6 aequorea ma
8	1075	84.4	238	Q8WPC5	Q8wp95 aequorea ma
9	1071	84.1	238	Q8WTC4	Q8wtc4 aequorea ma
10	1069	84.0	238	Q8WTD0	Q8wtc0 aequorea ma
11	1068	83.9	238	Q8WTC8	Q8wtc8 aequorea ma
12	1068	83.9	238	Q8WTC9	Q8wtc9 aequorea ma
13	1066	83.7	238	Q8WTC7	Q8wtc7 aequorea ma
14	1064	83.6	238	Q8WTC5	Q8wtc5 aequorea ma
15	251.5	19.8	225	Q95UA7	Q95ua7 montastraea
16	251.5	19.8	225	Q7Z0W5	Q7z0w5 montastraea

17	246	19.3	225	5	Q963F5	Q963f5 montastraea
18	243.5	19.1	225	5	Q7Z0W9	Q7z0w9 montastraea
19	243.5	19.1	236	5	Q8T6U0	Q8t6u0 dendronephc
20	239	18.8	225	5	Q8I6J8	Q8i6j8 trachyphyl1
21	237.5	18.7	266	5	Q9U6Y3	Q9u6y3 clavularia
22	233	18.3	225	5	Q8T5F1	Q8t5f1 montastraea
23	232	18.2	225	5	Q7Z0W4	Q7z0w4 montastraea
24	231	18.1	224	5	Q8MU48	Q8mu48 montastraea
25	215	16.9	227	5	Q7Z0W6	Q7z0w6 montastraea
26	215	16.9	234	5	Q7Z0W7	Q7z0w7 montastraea
27	211.5	16.6	259	5	Q8WMA2	Q8wma2 agaricia fr
28	211	16.6	239	5	Q8WMA1	Q8wma1 agaricia ag
29	210	16.5	234	5	Q8T5F2	Q8t5f2 montastraea
30	210	16.5	239	5	Q8MU47	Q8mu47 montastraea
31	209.5	16.5	229	5	Q9U6Y6	Q9u6y6 anemonia ma
32	207	16.3	227	5	Q962P9	Q962p9 montastraea
33	207	16.3	227	5	Q7Z0W8	Q7z0w8 montastraea
34	203	15.9	221	5	Q95P04	Q95p04 goniorpara t
35	203	15.9	227	5	Q95VTO	Q95vto montastraea
36	202.5	15.9	232	5	Q9GP15	Q9gpi5 anemonia su
37	202.5	15.9	238	5	Q9BLY9	Q9bly9 renilla mue
38	200.5	15.8	225	5	Q9U6Y8	Q9u6y8 discosoma s
39	198.5	15.6	214	5	Q86LV7	Q86lv7 meandrina m
40	198.5	15.6	232	5	Q9GZ28	Q9gz28 anemonia su
41	197.5	15.5	214	5	Q86LV8	Q86lv8 meandrina m
42	197.5	15.5	222	5	Q7Z168	Q7z168 cerianthus
43	197.5	15.5	225	5	Q8T6T9	Q8t6t9 radianthus
44	197	15.5	235	5	Q8T5F0	Q8t5f0 scolymia cu
45	196.5	15.4	232	5	Q9U6Y7	Q9u6y7 discosoma s

ALIGNMENTS

RESULT 1

Q8GHE2	ID	Q8GHE2	PRELIMINARY;	PRT;	238 AA.
AC	Q8GHE2;				
DT	01-MAR-2003 (TrEMBLrel. 23, Created)				
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)				
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)				
DE	Green fluorescence protein.				
GN	228GFP.				
OS	Azotobacter vinelandii.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;				
OC	Pseudomonadaceae; Azotobacter.				
OX	NCBI_TaxID=354;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=DSM2289;				
RA	Koranyi P., Berenyi M., Burg K.;				
RT	Occurrence of green fluorescence protein in diazotrophic bacteria				
RT	Azomonas and Azotobacter.;				
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF324408; AAN86140.1; -				
DR	GO; GO:0006091; P:energy pathways; IEA.				
DR	InterPro; IPR009017; GFP-like.				
DR	InterPro; IPR000786; Green_fl_protein.				
DR	Pfam; PF01353; GFP; 1.				
DR	PRINTS; PR01229; GFLUORESCENT.				
DR	ProDom; PD013756; Green_fl_protein; 1.				
SQ	SEQUENCE 238 AA; 26914 MW; F84840F1F9064018 CRC64;				

Query Match 97.2%; Score 1237; DB 2; Length 238;  
Best Local Similarity 97.5%; Pred. No. 1.5e-96;  
Matches 232; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy	2	VSKGEELFTGVVILVELDGVNGHKFSVSGEGDATYKGLTKLFCITCTGKLPVWPPTL	61
Db	1	MSKGEELFTGVVILVELDGVNGHKFSVSGEGDATYKGLTKLFCITCTGKLPVWPPTL	60
Qy	62	VTLLSYGVQCFSPYDPMKQHDFFKSMPEGYVQERTIPFKDGNKYKTRAEVFEGLTV	121

```
Db 61 VTTFSYGVCFSRYPDHMKRHFPSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
QY 122 NRIELKGIDFKEDGNILGHLEYNVYIMADKQNGIKVNFKIRHNIEDGSVQLAD 181
Db 121 NRIELKGIDFKEDGNILGHLEYNVYIMADKQNGIKVNFKIRHNIEDGSVQLAD 180
QY 182 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

RESULT 2
Q93125 ID Q93125 PRELIMINARY; PRT; 238 AA.
AC Q93125;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein mutant 3.
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96305137; PubMed=8707053;
RA Cormack B.P., Valdivia R.H., Falkow S.;
RT "FACS-optimized mutants of the green fluorescent protein (GFP).";
RL Gene 173:33-38(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Cormack B.P., Bertram G., Egerton M., Gow N.A.R., Falkow S.,
RA Brown A.J.P.;
RT "Yeast Enhanced Green Fluorescent Protein (yEGFP): a reporter of gene
RT expression in Candida albicans.";
RL Microbiology 0:0-0(1996).
DR EMBL; U73901; AAB18957.1; -.
DR HSSP; P42212; 1BFP.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR PRODOM; PD013756; Green_fl_protein; 1.
DR SEQUENCE 238 AA; 26840 MW; A28622809A9DEA60 CRC64;

Query Match 96.9%; Score 1234; DB 5; Length 238;
Best Local Similarity 97.1%; Pred. No. 2.7e-96;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPILVELDGVNNGHKFSVSGEGDATYKGLTKLFTCTTGKLPVWPPTL 61
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSKGEELFTGVVPILVELDGVNNGHKFSVSGEGDATYKGLTKLFTCTTGKLPVWPPTL 60

QY 52 VTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 121
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 51 VTTFGYGVQCFARYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHLEYNVYIMADKQNGIKVNFKIRHNIEDGSVQLAD 181
Db 121 NRIELKGIDFKEDGNILGHLEYNVYIMADKQNGIKVNFKIRHNIEDGSVQLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

RESULT 3
Q8GHE4 ID Q8GHE4 PRELIMINARY; PRT; 238 AA.
AC Q8GHE4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Green fluorescence protein.
GN 85GFP.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
RT Azomonas and Azotobacter.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324406; AAN86138.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR PRODOM; PD013756; Green_fl_protein; 1.
DR SEQUENCE 238 AA; 26887 MW; E0E1616BD2AF6188 CRC64;

Query Match 96.7%; Score 1231; DB 2; Length 238;
```

```
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescence protein.
GN 375GFP.
OS Azomonas agilis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azomonas.
OX NCBI_TaxID=116849;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
RT Azomonas and Azotobacter.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324405; AAN86137.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR PRODOM; PD013756; Green_fl_protein; 1.
DR SEQUENCE 238 AA; 26902 MW; 15FE9B9C5B4F6B89 CRC64;

Query Match 96.8%; Score 1232; DB 2; Length 238;
Best Local Similarity 97.1%; Pred. No. 4e-96;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPILVELDGVNNGHKFSVSGEGDATYKGLTKLFTCTTGKLPVWPPTL 61
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGVVPILVELDGVNNGHKFSVSGEGDATYKGLTKLFTCTTGKLPVWPPTL 60

QY 62 VTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 121
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTFSYGVCFSRYPDHMKRHFPSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHLEYNVYIMADKQNGIKVNFKIRHNIEDGSVQLAD 181
Db 121 NRIELKGIDFKEDGNILGHLEYNVYIMADKQNGIKVNFKIRHNIEDGSVQLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

RESULT 4
Q8GHE3 ID Q8GHE3 PRELIMINARY; PRT; 238 AA.
AC Q8GHE3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescence protein.
GN 85GFP.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
RT Azomonas and Azotobacter.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324406; AAN86138.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR PRODOM; PD013756; Green_fl_protein; 1.
DR SEQUENCE 238 AA; 26887 MW; E0E1616BD2AF6188 CRC64;

Query Match 96.7%; Score 1231; DB 2; Length 238;
```

```
Best Local Similarity 97.1%; Pred. No. 4.9e-96;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 60
QY 62 VTLLSYGVQCFSRYPDHMKOHDFKSPAMPEGVYQERTIFFKDDGNKTRAEVKFEGDTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKRHDFKSPAMPEGVYQERTIFFKDDGNKTRAEVKFEGDTLV 120
QY 122 NRIELKGIDFKEDGNILGHLENYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHLENYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180
QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLLVGFTVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLLVGFTVTAAGITLGMDELYK 238

RESULT 5
Q17105
ID Q17105 PRELIMINARY; PRT; 238 AA.
AC Q17105;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83959; CAA58789.1; -.
DR PIR; JS0692; JQ1514.
DR HSSP; P42212; 1GFL.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748E44 CRC64;

Query Match 94.2%; Score 1199; DB 5; Length 238;
Best Local Similarity 93.7%; Pred. No. 2.5e-93;
Matches 223; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 60
QY 62 VTLLSYGVQCFSRYPDHMKOHDFKSPAMPEGVYQERTIFFKDDGNKTRAEVKFEGDTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKOHDFKSPAMPEGVYQERTIFFKDDGNKTRAEVKFEGDTLV 120
QY 122 NRIELKGIDFKEDGNILGHLENYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHLENYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180
QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLLVGFTVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLLVGFTVTAAGITLGMDELYK 238

RESULT 6
Q17106
ID Q17106 PRELIMINARY; PRT; 238 AA.
AC Q17106;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83959; CAA58789.1; -.
DR PIR; JS0692; JQ1514.
DR HSSP; P42212; 1GFL.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748E44 CRC64;

Query Match 93.0%; Score 1184; DB 5; Length 238;
Best Local Similarity 92.9%; Pred. No. 4.6e-92;
Matches 221; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 60
QY 62 VTLLSYGVQCFSRYPDHMKOHDFKSPAMPEGVYQERTIFFKDDGNKTRAEVKFEGDTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKOHDFKSPAMPEGVYQERTIFFKDDGNKTRAEVKFEGDTLV 120
QY 122 NRIELKGIDFKEDGNILGHLENYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHLENYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180
QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLLVGFTVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLLVGFTVTAAGITLGMDELYK 238

RESULT 7
Q8WTC6
ID Q8WTC6 PRELIMINARY; PRT; 238 AA.
AC Q8WTC6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPxm19uv;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.O., Chen M.,
RA Li S.J., Xia N.S.;
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435431; AAL33916.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP_like.
DR ProDom; PD013756; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
```

```
DR PRINTS; PR01229; GFP; 238 AA; 27015 MW; 688FD75E88926903 CRC64;
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27015 MW; 688FD75E88926903 CRC64;

Query Match 84.8%; Score 1079; DB 5; Length 238;
Best Local Similarity 81.9%; Pred. No. 3.4e-83;
Matches 195; Conservative 20; Mismatches 23; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDGYKLTILKFICTTGKLPVWPPTL 61
Db 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDGYKLTILKFICTTGKLPVWPPTL 60

QY 62 VTLLSYGVQCFSPYDPHMKOHDFPKSAMPEGVYQERTIFFKDDGNKYKTRAEVKEGDTLV 121
Db 61 VTLLSYGVQCFSPYDPHMKOHDFPKSAMPEGVYQERTIFFKDDGNKYKTRAEVKEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYVIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
Db 121 NRIELKGIDFKEDGNILGHKLEYNHNHYVIMADKQNGIKVNFKIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 238

RESULT 8
Q8WP95 PRELIMINARY; PRT; 238 AA.
AC Q8WP95;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFPXM.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL33918.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000917; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match 84.1%; Score 1071; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 1.6e-82;
Matches 196; Conservative 17; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDGYKLTILKFICTTGKLPVWPPTL 61
Db 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDGYKLTILKFICTTGKLPVWPPTL 60

QY 62 VTLLSYGVQCFSPYDPHMKOHDFPKSAMPEGVYQERTIFFKDDGNKYKTRAEVKEGDTLV 121
Db 61 VTLLSYGVQCFSPYDPHMKOHDFPKSAMPEGVYQERTIFFKDDGNKYKTRAEVKEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYVIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
Db 121 NRIELKGIDFKEDGNILGHKLEYNHNHYVIMADKQNGIKVNFKIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 238

RESULT 9
Q8WTC4 PRELIMINARY; PRT; 238 AA.
AC Q8WTC4;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL33918.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000917; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match 84.1%; Score 1071; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 1.6e-82;
Matches 196; Conservative 17; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDGYKLTILKFICTTGKLPVWPPTL 61
Db 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDGYKLTILKFICTTGKLPVWPPTL 60

QY 62 VTLLSYGVQCFSPYDPHMKOHDFPKSAMPEGVYQERTIFFKDDGNKYKTRAEVKEGDTLV 121
Db 61 VTLLSYGVQCFSPYDPHMKOHDFPKSAMPEGVYQERTIFFKDDGNKYKTRAEVKEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYVIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
Db 121 NRIELKGIDFKEDGNILGHKLEYNHNHYVIMADKQNGIKVNFKIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 238

RESULT 10
Q8WTD0 PRELIMINARY; PRT; 238 AA.
AC Q8WTD0;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL33918.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000917; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match 84.8%; Score 1075; DB 5; Length 238;
Best Local Similarity 81.5%; Pred. No. 7.4e-83;
Matches 194; Conservative 20; Mismatches 24; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDGYKLTILKFICTTGKLPVWPPTL 61
Db 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQSGEGDGYKLTILKFICTTGKLPVWPPTL 60

QY 62 VTLLSYGVQCFSPYDPHMKOHDFPKSAMPEGVYQERTIFFKDDGNKYKTRAEVKEGDTLV 121
Db 61 VTLLSYGVQCFSPYDPHMKOHDFPKSAMPEGVYQERTIFFKDDGNKYKTRAEVKEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYVIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
Db 121 NRIELKGIDFKEDGNILGHKLEYNHNHYVIMADKQNGIKVNFKIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 238
```

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF435427; AAL33912.1; -.  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR0009017; GFP like.  
DR InterPro; IPR000786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFLUORESCENT.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
SQ SEQUENCE 238 AA; 26997 MW; 5F80A192173CB84D CRC64;

Query Match 84.0%; Score 1069; DB 5; Length 238;  
Best Local Similarity 81.1%; Pred. No. 2.4e-82;  
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPIVLVDGNGHFKFSVSGEGDATYKGLTKLFCITGKLPVPWPTL 61  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
1 MSKGELFTGIVPVLIELDGDVHGKFSVRGEGDADYKLEIKFICTTGKLPVPWPTL 60  
QY 62 VTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
61 VTTLGVIQCFARYPEHMKQNDFFKSAMPEGYIERTIFFQDDGKYKTRGEVKFEGDTLV 120  
QY 122 NRLEKGDIDFKEDGNILGHKLEYNYNHNHVMADKQKNGIKVNFKIRHNIEDGSVOLAD 181  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
121 NRLEKGMDFKEDGNILGHKLEYNFNHNVIMPDKANNGLKVNFKIRHNIEGGVOLAD 180  
QY 182 HYQONTPIGDGPVLLPDNHYLSQTALSQSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
181 HYQTNVPLGDGPVLLPINHYLSQTAISKDRNETRDHVMVLFEEFSACGHTGMDLYK 238

RESULT 11

Q8WTC8 PRELIMINARY; PRT; 238 AA.

AC Q8WTC8;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Green fluorescent protein.  
GN GFP.  
OS Aequorea macrodactyla.  
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;  
OC Aequoreidae; Aequorea.  
OX NCBI\_TaxID=147615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GFPxm163;  
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,  
RA Li S.J., Xia N.S.;  
RT "Colorful mutants of green fluorescent protein from Aequorea  
macrodactyla.";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF435429; AAL33914.1; -.  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR0009017; GFP like.  
DR InterPro; IPR000786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFLUORESCENT.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;

Query Match 83.9%; Score 1068; DB 5; Length 238;  
Best Local Similarity 81.1%; Pred. No. 2.9e-82;  
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPIVLVDGNGHFKFSVSGEGDATYKGLTKLFCITGKLPVPWPTL 61  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
1 MSKGELFTGIVPVLIELDGDVHGKFSVRGEGDADYKLEIKFICTTGKLPVPWPTL 60  
QY 62 VTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
61 VTTLGVIQCFARYPEHMKQNDFFKSAMPEGYIERTIFFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRLEKGDIDFKEDGNILGHKLEYNYNHNHVMADKQKNGIKVNFKIRHNIEDGSVOLAD 181  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
121 NRLEKGMDFKEDGNILGHKLEYNFNHNVIMPDKANNGLKVNFKIRHNIEGGVOLAD 180  
QY 182 HYQONTPIGDGPVLLPDNHYLSQTALSQSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
181 HYQTNVPLGDGPVLLPINHYLSQTAISKDRNETRDHVMVLFEEFSACGHTGMDLYK 238

RESULT 12

Q8WTC9 PRELIMINARY; PRT; 238 AA.

AC Q8WTC9;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Green fluorescent protein.  
GN GFP.  
OS Aequorea macrodactyla.  
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;  
OC Aequoreidae; Aequorea.  
OX NCBI\_TaxID=147615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GFPxm162;  
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,  
RA Li S.J., Xia N.S.;  
RT "Colorful mutants of green fluorescent protein from Aequorea  
macrodactyla.";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF435428; AAL33913.1; -.  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR0009017; GFP like.  
DR InterPro; IPR000786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFLUORESCENT.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
SQ SEQUENCE 238 AA; 27031 MW; 5F80A19C19DC584D CRC64;

Query Match 83.9%; Score 1068; DB 5; Length 238;  
Best Local Similarity 81.1%; Pred. No. 2.9e-82;  
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPIVLVDGNGHFKFSVSGEGDATYKGLTKLFCITGKLPVPWPTL 61  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
1 MSKGELFTGIVPVLIELDGDVHGKFSVRGEGDADYKLEIKFICTTGKLPVPWPTL 60  
QY 62 VTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
61 VTTLGVIQCFARYPEHMKQNDFFKSAMPEGYIERTIFFQDDGKYKTRGEVKFEGDTLV 120  
QY 122 NRLEKGDIDFKEDGNILGHKLEYNYNHNHVMADKQKNGIKVNFKIRHNIEDGSVOLAD 181  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
121 NRLEKGMDFKEDGNILGHKLEYNFNHNVIMPDKANNGLKVNFKIRHNIEGGVOLAD 180  
QY 182 HYQONTPIGDGPVLLPDNHYLSQTALSQSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
181 HYQTNVPLGDGPVLLPINHYLSQTAISKDRNETRDHVMVLFEEFSACGHTGMDLYK 238

RESULT 13

Q8WTC7 PRELIMINARY; PRT; 238 AA.

AC Q8WTC7;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Green fluorescent protein.  
GN GFP.  
OS Aequorea macrodactyla.  
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;  
OC Aequoreidae; Aequorea.  
OX NCBI\_TaxID=147615;

Query Match 83.68; Score 1064; DB 5; Length 238;  
Best Local Similarity 81.9%; Pred. No. 6.3e-8;  
Matches 195; Conservative 26; Indels 0; Gaps 0;  
2 VSKGRFPTGVVPTLVELDGMVGHKFSVSGEGDQTYGKLTLKFICTTGKLPVPTL 61



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:41:49 ; Search time 47.1111 Seconds  
(without alignments)  
1433.395 Million cell updates/sec

Title: US-09-887-784-64V

Perfect score: 1273  
Sequence: 1 MYSGBELFTGVVILVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: Genesecp1980s:\*
- 2: Genesecp1990s:\*
- 3: Genesecp2000s:\*
- 4: Genesecp2001s:\*
- 5: Genesecp2002s:\*
- 6: Genesecp2003as:\*
- 7: Genesecp2003bs:\*
- 8: Genesecp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1269	99.7	239	5	Aae17518 Enhanced
2	1269	99.7	363	6	AbR40352 Human ami
3	1269	99.7	893	4	AgG5781 Amino aci
4	1269	99.7	1132	4	AgG5782 Amino aci
5	1261	99.1	239	5	Aae17517 Enhanced
6	1258	98.8	239	3	Aab22882 Enhanced
7	1258	98.8	239	3	Aay54349 Amino aci
8	1258	98.8	239	3	Aay79584 EGFP sign
9	1258	98.8	239	4	Aab50804 Jellyfish
10	1258	98.8	239	4	Aab85900 A. victor
11	1258	98.8	239	4	Aab31171 Amino aci
12	1258	98.8	239	5	AgG66198 A. victor
13	1258	98.8	239	5	AgG94444 Protease
14	1258	98.8	239	5	Aae14599 Aequorea
15	1258	98.8	239	6	Aae34958 Aequorea
16	1258	98.8	239	6	AgG79829 Green flu
17	1258	98.8	239	6	AbR83616 Green flu
18	1258	98.8	239	6	Ada38074 Aequorea
19	1258	98.8	239	7	AbU63204 Aequorea
20	1258	98.8	239	7	Adc18358 EGFP (enh
21	1258	98.8	239	7	AbW00914 Aequorea
22	1258	98.8	239	7	AdE28570 Enhanced
23	1258	98.8	246	7	Abm79011 Enhanced
24	1258	98.8	248	5	AgG68319 Jellyfish
25	1258	98.8	259	5	Aau99804 Biomembra

26	1258	98.8	265	2	AAW97451	Aaw97451 Wild-type
27	1258	98.8	268	5	AAU99803	Aau99803 Biomembra
28	1258	98.8	270	5	AAU99802	Aau99802 Biomembra
29	1258	98.8	272	5	AAU99800	Aau99800 Biomembra
30	1258	98.8	273	5	AAU99801	Aau99801 Biomembra
31	1258	98.8	280	5	AAU99807	Aau99807 Biomembra
32	1258	98.8	281	3	AAU50142	Aay50142 Green flu
33	1258	98.8	281	3	AAU50142	Aab24252 EGFP-MODC
34	1258	98.8	281	5	AAU10888	Aau10888 EGFP-MODC
35	1258	98.8	286	7	AdE28562	AdE28562 EGFP/ hum
36	1258	98.8	289	7	AdE28564	AdE28564 EGFP/ hum
37	1258	98.8	290	7	AdE28568	AdE28568 EGFP/ hum
38	1258	98.8	290	7	AdE28566	AdE28566 EGFP/ hum
39	1258	98.8	294	3	AAU322860	Aab22860 GFP-DSVD-
40	1258	98.8	294	3	AAU79638	Aay79638 Caspase-3
41	1258	98.8	294	5	ABG94422	ABG94422 Recombina
42	1258	98.8	308	2	AAU42181	Aay42181 EGFP/DRM
43	1258	98.8	320	6	ABR83620	ABR83620 HUB1-GFP
44	1258	98.8	323	3	AAU54359	Aay54359 GFP mutan
45	1258	98.8	323	6	ABR83621	ABR83621 RUB1-GFP

ALIGNMENTS

RESULT 1  
AAE17518  
ID AAE17518 standard; protein; 239 AA.  
XX  
AC AAE17518;  
DT  
DT 22-APR-2002 (first entry)  
XX  
DE Enhanced F64L-E222G jellyfish green fluorescent protein mutant.  
XX  
KW Jellyfish; green fluorescent protein; GFP; protein redistribution;  
KW cellular function; genetic reporter; mutant; Stoke's shift; muten.  
XX  
OS Aequorea victoria.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 65 /note= "Wild type Phe substituted with Leu; This  
FT corresponds to position 64 in the wild type protein"  
FT  
FT Misc-difference 223 /note= "Wild type Glu substituted with Gly; This  
FT corresponds to position 222 in the wild type protein"  
FT

WO200198338-A2.  
27-DEC-2001.  
18-JUN-2001; 2001WO-EP006848.  
19-JUN-2000; 2000DK-00000953.  
20-JUN-2000; 2000US-0212681P.  
10-MAY-2001; 2001DK-00000739.  
10-MAY-2001; 2001US-0290170P.  
(BIOI-) BIOIMAGE AS.

Bjorn SP, Pagliaro L, Thastrup O;  
WPI; 2002-098224/13.  
DR N-PSDB; AAD28163.  
XX  
PT Novel fluorescent protein in vitro assay for measuring protein kinase  
PT activity or dephosphorylation activity, or for measuring protein  
PT redistribution, has a green fluorescent protein with F64L and E222G  
PT mutation.  
XX  
PS Claim 9; Page 37; 41pp; English.

XX CC The invention relates to a fluorescent protein derived from green  
CC fluorescent protein (GFP) or its analogue. The GFP containing mutations  
CC at F64L and E222G has a bigger compared to other GFP's making it very  
CC suitable for high throughput screening due to better resolution. The  
CC fluorescent protein is useful in invitro assays for measuring protein  
CC kinase activity or dephosphorylation activity, or for measuring protein  
CC redistribution. The fluorescent protein is useful in studying cellular  
CC functions in living cells; as protein tags in transgenic animals, living  
CC and fixed cells; organelle tags, secretion marker and genetic reporter.  
CC The fluorescent protein is also useful as a cell or organelle integrity  
CC marker, a marker for changes in cell morphology, as transfection marker,  
CC and as a marker to be used in combination with fluorescence activated  
CC cell sorting (FACS). The novel proteins can also be used as reporters to  
CC monitor live or dead biomass of organisms, such as fungi. The fluorescent  
CC protein is also useful as markers in transcriptional and translational  
CC fusions for performing transposon vector mutagenesis and as a reporter  
CC for bacterial detection. Transposons encoding the fluorescent protein are  
CC useful for screening promoters and for tagging plasmids and chromosomes.  
CC The fluorescent protein engineered into the genome of a phage is useful  
CC for designing diagnostic tool. The present sequence is a DNA encoding  
CC enhanced F64L-E222G jellyfish green fluorescent protein (GFP) mutant  
XX SQ Sequence 239 AA;

Query Match 99.7%; Score 1269; DB 5; Length 239;  
Best Local Similarity 99.6%; Pred. No. 9.7e-123;  
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDNGHKFSVSGEGGDATYKGLTLKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDNGHKFSVSGEGGDATYKGLTLKFICTTGKLPVPWPT 60  
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 2  
ABR40352  
ID ABR40352 standard; protein; 363 AA.  
XX ABR40352;  
XX  
XX  
XX 08-JUL-2003 (first entry)  
XX  
XX Human amino acid sequence SEQ ID NO: 6.  
XX Human; heterologous conjugate; intracellular protein.  
XX Homo sapiens.  
XX Aequoria victoria.  
XX WO2003029827-A2.  
XX  
XX 10-APR-2003.  
XX  
XX 01-OCT-2002; 2002WO-DK000651.  
XX PF  
XX 01-OCT-2001; 2001DK-00001433.  
XX PR  
XX 11-OCT-2001; 2001US-0328896P.  
XX  
XX (BIOL-) BIOLMAGE AS.  
XX PA  
XX Terry BR, Nielsen SJ;  
XX PI

XX DR WPI: 2003-430211/40.  
DR N-PSDB; ACC72604.  
XX  
XX Novel cell for identifying modulators of protein interaction, contains a  
PT first conjugate comprising anchor protein, second conjugate having type B  
PT interactor protein and third conjugate with detectable group.  
XX  
XX Disclosure; Page 112-113; 118pp; English.  
XX  
XX The invention relates to a novel cell, comprising three heterologous  
CC conjugates (HC), a first HC (HC1) comprising an anchor protein that  
CC specifically binds to an internal structure within the cell conjugated to  
CC an interactor protein (IP) of type A, a second HC (HC2) comprising IP of  
CC type B conjugated to a first protein of interest, and a third HC (HC3)  
CC comprising a second protein of interest conjugated to detectable group.  
CC The cell is useful for detecting if a compound disrupts or induces the  
CC interaction between two intracellular proteins. The cell is also useful  
CC for screening compounds that modulate the interaction between two  
CC intracellular proteins. The present sequence is used in the  
CC exemplification of the invention  
XX SQ Sequence 363 AA;

Query Match 99.7%; Score 1269; DB 6; Length 363;  
Best Local Similarity 99.6%; Pred. No. 1.8e-122;  
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDNGHKFSVSGEGGDATYKGLTLKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDNGHKFSVSGEGGDATYKGLTLKFICTTGKLPVPWPT 60  
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 3  
AAG65781  
ID AAG65781 standard; protein; 893 AA.  
XX AAG65781;  
XX  
XX 07-JAN-2002 (first entry)  
XX  
XX Amino acid sequence of HSPDB4A1-E222G fusion protein.  
XX  
XX pDE4; central nervous system; antiinflammatory; cytostatic; nootropic;  
XX autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;  
XX fusion protein.  
XX Homo sapiens.  
XX Aequorea victoria.  
XX WO200179526-A2.  
XX  
XX 25-OCT-2001.  
XX  
XX 11-APR-2001; 2001WO-DK000264.  
XX  
XX 17-APR-2000; 2000DK-00000651.  
XX PR  
XX 29-MAY-2000; 2000DK-00000849.  
XX  
XX (BIOI-) BIOIMAGE AS.  
XX PA  
XX  
XX

PI Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;  
PI Praestegaard M;  
XX  
XX  
DR WPI; 2001-611727/70.  
DR N-PSDB; AAI66852.  
XX  
XX  
PT Determining if a compound is a dislocator of PDE4 for identifying  
PT compounds for treating CNS and inflammatory disease comprises identifying  
PT compounds which remove PDE4 spots.  
XX  
XX  
XX Example 1; Page 156-160; 160pp; English.  
XX  
XX The invention relates to determining, if a compound, is a dislocator of  
XX PDE4. The method comprises testing if the compound removes PDE4 spots,  
XX which may optionally be induced by a Rolipram-like reference compound,  
XX and testing if it inhibits the catalytic activity of the PDE4, where the  
XX compound is a dislocator of PDE4, if it removes PDE spots and if it does  
XX not inhibit the catalytic activity of PDE4. The method is useful for  
XX identifying compounds useful for the treatment of diseases of the central  
XX nervous system such as depression and for the treatment of inflammatory  
XX disease such as joint inflammation, Crohn's disease, inflammatory bowel  
XX disease, respiratory diseases, chronic obstructive pulmonary disease  
XX (COPD), including asthma, chronic bronchitis, pulmonary emphysema,  
XX endotoxin shock, toxic shock syndrome, systemic lupus erythematosus,  
XX psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV  
XX infection. The use of a reagent that can mimic or reverse the effect of  
XX the compound with affinity for the catalytic site on intracellular  
XX distribution of the PDE for the preparation of a medicament. The present  
XX sequence represents the amino acid sequence of a HSPDE4A1-E222G fusion  
XX protein  
XX  
XX Sequence 893 AA;  
XX  
XX Query Match 99.7%; Score 1269; DB 4; Length 893;  
XX Best Local Similarity 99.6%; Pred. No. 6.9e-122;  
XX Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYKGLTLKFTCTTGKLPVPMPT 60  
DB 655 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYKGLTLKFTCTTGKLPVPMPT 714  
QY 61 LVTVLISYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNKYKTRAEVKEGDTL 120  
DB 715 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNKYKTRAEVKEGDTL 774  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180  
DB 775 VNRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKVNFKIRHNIEDGSVOLA 834  
QY 181 DHYQQNTPIGDGVPVLLPDNHYLSTQSALSADKPNKRDRHMLLGFVTAAGITLGMDELYK 239  
DB 835 DHYQQNTPIGDGVPVLLPDNHYLSTQSALSADKPNKRDRHMLLGFVTAAGITLGMDELYK 893  
XX  
XX RESULT 4  
XX AAG65782  
XX ID AAG65782 standard; protein; 1132 AA.  
XX  
XX AC AAG65782;  
XX  
XX XX 07-JAN-2002 (first entry)  
XX  
XX DE Amino acid sequence of HSPDE4A4-E222G fusion protein.  
XX  
XX XX PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;  
XX KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;  
XX fusion protein.  
XX  
XX XX Homo sapiens.  
XX OS Aquorea victoria.  
XX  
XX XX WO200179526-A2.  
XX  
XX



CC biosensors of the invention can be used to investigate a wide range of  
CC cellular activities and to screen compounds which modulate these  
CC activities. Biosensors containing a recognition site for caspase, for  
CC example, may be used for the screening of compounds which modulate  
CC apoptosis, while biosensors containing other protease recognition sites  
CC may be used for the detection of proteolytic toxins (such as anthrax  
CC lethal factor). The method provides improved target validation and  
CC candidate compound optimisation by combining many cell screening formats  
CC with fluorescence-based molecular reagents and computer-based feature  
CC extraction, data analysis and automation, resulting in increased quantity  
CC and speed of data collection and faster evaluation of drug candidates.  
CC Sequences AAB2881-B2885 represent fluorescent proteins which may be used  
CC as components of biosensor fusion proteins of the invention  
XX  
SQ Sequence 239 AA;

Query Match 98.8%; Score 1258; DB 3; Length 239;  
Best Local Similarity 98.7%; Pred. No. 1.3e-121;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYKLTAKFICTTGKLPVWPWT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYKLTAKFICTTGKLPVWPWT 60  
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
QY 181 DHVQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 239  
DB 181 DHVQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 239

RESULT 7  
AA54349  
ID AAY54349 standard; protein; 239 AA.  
XX  
AC AAY54349;  
DT 06-APR-2000 (first entry)  
XX  
DE Amino acid sequence of the mutant green fluorescent protein EGFP.  
XX  
KW Fluorescent protein; green fluorescent protein; emission intensity;  
KW fluorescence; pH detection; pH sensor; EGFP.  
XX  
OS Synthetic.  
OS Aequorea victoria.

Key Location/Qualifiers  
FH Misc-difference 65 /note= "wild type Phe substituted with Leu"  
FT Misc-difference 66 /note= "wild type Ser substituted with Thr"  
FT Misc-difference 232 /note= "wild type His substituted with Leu"  
XX  
XX WO964592-A2.  
XX  
XX 16-DEC-1999.  
XX  
XX 08-JUN-1999; 99WO-US012850.  
XX  
XX 09-JUN-1998; 98US-00094359.  
XX 13-OCT-1998; 98US-00172063.  
XX  
XX (REGC) UNIV CALIFORNIA.  
XX (UYOR-) UNIV OREGON STATE.  
XX

PI Tsien RY, Llopis J, Wachter RM;  
XX WPI; 2000-116540/10.  
DR N-PSDB; AA245642.  
XX  
PT New functional engineered green fluorescent proteins, used for measuring  
PT the pH in biological samples and cells.  
XX  
PS Disclosure; Page 9; 89pp; English.  
XX  
CC The present sequence represents a functional engineered fluorescent  
CC protein based on the Aequorea green fluorescent protein (GFP). The  
CC emission intensity changes as pH varies between 5 and 10 of the present  
CC protein are novel. The functional engineered fluorescent proteins show  
CC reversible changes in fluorescence over physiological pH ranges. They can  
CC be used for determining the pH of samples and cells. The polynucleotides  
CC can also be used to produce transgenic animals. The fluorescent protein  
CC pH sensors can be delivered to cells in the form of polynucleotides  
CC encoding the protein sensor fused to a targeting signal. The targeting  
CC signal directs the expression of the protein sensors to restricted cell  
CC locations. This makes it possible to measure the pH of a precisely  
CC defined cellular region or organelle  
XX  
SQ Sequence 239 AA;

Query Match 98.8%; Score 1258; DB 3; Length 239;  
Best Local Similarity 98.7%; Pred. No. 1.3e-121;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYKLTAKFICTTGKLPVWPWT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYKLTAKFICTTGKLPVWPWT 60  
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
QY 181 DHVQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 239  
DB 181 DHVQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 239

RESULT 8  
AA54349  
ID AAY79584 standard; peptide; 239 AA.  
XX  
AC AAY79584;  
DT 29-AUG-2000 (first entry)  
XX  
DE EGFP signal domain.  
XX  
KW Protease; biosensor; EGFP; signal peptide; cell screening; assay;  
KW analysis; drug discovery.  
XX  
OS Unidentified.  
XX  
XX WO200026408-A2.  
XX  
XX 11-MAY-2000.  
XX  
XX 29-OCT-1999; 99WO-US025431.  
XX  
XX 30-OCT-1998; 98US-0106308P.  
XX 26-MAY-1999; 99US-0136078P.  
XX  
XX (CELL-) CELLOMICS INC.  
XX  
XX Guiliano KA, Bright G, Olson K, Burroughs-Tencza S;  
PI

XX WPI; 2000-365644/31.  
 DR N-PSDB; AAA27573.  
 XX Recombinant nucleic acid encoding a protease biosensor useful for  
 PT fluorescence based cell and molecular biochemical assays for drug  
 PT discovery comprising three operably linked nucleic acid sequences.  
 XX  
 XX Claim 14; Fig 29A; 218pp; English.  
 XX  
 CC The present sequence is that of the EGFP signal domain, which can be  
 CC included in novel recombinant protease biosensors (PBs) of the invention.  
 CC The PBs (see AAY79638-54) comprise: a first domain (see AAY79579-87)  
 CC comprising at least 1 detectable polypeptide signal such as the present  
 CC sequence; a second domain (see AAY79588-622) comprising at least 1  
 CC protease recognition site; and a third domain (see AAY79623-37)  
 CC comprising at least 1 reactant target sequence. A recombinant nucleic  
 CC acid (see AAA27627-43) encoding the PB, an expression vector, and a  
 CC genetically engineered host cell are also claimed. A claimed method for  
 CC identifying compounds that modify protease activity in a cell involves  
 CC contacting a host cell that possesses the recombinant PB with a test  
 CC compound, and determining the PB distribution in the host cell, where  
 CC changes in the distribution of the PB are correlated with modification of  
 CC protease activity by the test compound. Claimed kits for identifying  
 CC compounds that modify protease activity in a host cell include the  
 CC recombinant nucleic acid, or the recombinant PB, or the vector, or the  
 CC host cell. The PB is useful in high content screens to detect in vivo  
 CC activation of enzymatic activity, and to identify specific activity based  
 CC on cleavage of a known recognition motif  
 XX  
 SQ Sequence 239 AA;

Query Match 98.8%; Score 1258; DB 3; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 1.3e-121;  
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60  
 Db 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60  
 QY 61 LVTVLVSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 Db 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 Db 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 QY 131 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
 Db 131 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 9  
 AAB50804  
 ID AAB50804 standard; protein; 239 AA.  
 XX  
 AC AAB50804;  
 XX  
 DT 14-MAR-2001 (first entry)  
 XX  
 DE Jellyfish GFP mutant EGFP.  
 XX  
 KW Aequorea victoria; jellyfish; fluorescent protein indicator;  
 KW green fluorescent protein; GFP; linker moiety; sensor;  
 KW calmodulin-binding domain; mutant; mutein.  
 XX  
 OS Aequorea victoria.  
 XX  
 PN WO2000/71565-A2.  
 XX  
 PD 30-NOV-2000.  
 XX

PF 17-MAY-2000; 2000WO-US013684.  
 XX  
 PR 21-MAY-1999; 99US-00316919.  
 PR 21-MAY-1999; 99US-00316920.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 XX Tsien RV, Baird GA;  
 XX WPI; 2001-032017/04.  
 DR N-PSDB; AAC90488.  
 XX  
 PT Novel fluorescent proteins comprising a sensor protein inserted into  
 PT them, useful for measuring the response of a sensor biological, chemical,  
 PT electrical or physiological parameter in vivo or in vitro.  
 XX  
 PS Disclosure; Page 24; 94pp; English.  
 XX  
 CC The present sequence is a fluorescent protein used in the construction of  
 CC a fluorescent protein indicator. The indicator comprises a sensor  
 CC polypeptide that is responsive to a chemical, biological, electrical or  
 CC physiological parameter, and a fluorescence protein functional group. The  
 CC sensor polypeptide is operatively inserted into the fluorescent moiety.  
 CC The fluorescent indicator is useful for detecting the presence of a  
 CC response inducing member in a sample. The method involves contacting the  
 CC sample with the indicator and detecting a change in fluorescence, in  
 CC which a change is indicative of the effect of the parameter on the sensor  
 CC polypeptide. The novel fluorescent proteins are advantageous due to their  
 CC reduced size as compared to the FRET (fluorescence resonance energy  
 CC transfer)-based sensors  
 XX  
 SQ Sequence 239 AA;

Query Match 98.8%; Score 1258; DB 4; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 1.3e-121;  
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60  
 Db 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60  
 QY 61 LVTVLVSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 Db 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 Db 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
 Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 10  
 AAB85900  
 ID AAB85900 standard; protein; 239 AA.  
 XX  
 AC AAB85900;  
 XX  
 DT 30-NOV-2001 (first entry)  
 XX  
 DE A. victoria green fluorescent protein (GFP) and linker sequence.  
 XX  
 KW Melanin concentrating hormone receptor; MCH; chimeric; fusion;  
 KW fluorescent polypeptide; orexigenic; anabolic; food intake; GFP;  
 KW green fluorescent protein.  
 XX  
 OS Synthetic.  
 OS Aequorea victoria.  
 XX  
 PN WO200168706-A1.  
 XX

```

PD 20-SEP-2001.
XX
PF 14-MAR-2001; 2001WO-US008071.
XX
PR 15-MAR-2000; 2000US-0189698P.
XX
PA (MERI ) MERCK & CO INC.
XX
PI Marsh DJ;
XX
XX WPI; 2001-565791/63.
DR N-PSDB; AAH47304.
XX
XX Fusion proteins comprising melanin concentrating hormone receptor
PT peptides and fluorescent proteins, useful for identifying appetite
PT stimulants.
XX
XX Claim 2; Page 14; 71pp; English.
XX
XX The invention provides melanin concentrating hormone (MCH) receptor
CC (MCHR) chimeric and fusion proteins. The MCHR chimeric proteins comprise
CC MCHR polypeptide regions from different species. The MCHR fusion protein
CC comprise MCHR polypeptide region and a fluorescent polypeptide region
CC joined directly, or via a linker, to the carboxy side of the MCHR
CC polypeptide region. The MCHR fusion proteins can be expressed by standard
CC recombinant methodology. MCH action promotes feeding (orexigenic) and up
CC regulation of MCH activity stimulates food intake. The present sequence
CC represents a A. victoria green fluorescent protein (GFP) and a linker
CC sequence
XX
SQ Sequence 239 AA;

Query Match 98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.3e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTILKFICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTILKFICTTGKLPVWPWT 60
QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

RESULT 11
AAB31171
ID AAB31171 standard; protein; 239 AA.
XX
AC AAB31171;
XX
DT 02-APR-2001 (first entry)
XX
XX Amino acid sequence of a green fluorescent protein (GFP).
DE
XX Growth rate; death rate; reporter gene; luminescent protein;
KW fluorescent product; luciferase; green fluorescent protein; GFP.
XX
OS Aequorea victoria.
XX
PN WO200075367-A1.
XX
PD 14-DEC-2000.
XX
PF 07-JUN-2000; 2000WO-FI000507.

XX
PR 07-JUN-1999; 99FI-00001296.
PA (LILI/) LILIUS E.
PA (VIRT/) VIRTA M.
XX
PI Lilius E, Virta M;
XX
DR WPI; 2001-061737/07.
DR N-PSDB; AAC86954.
XX
XX Assessing growth and death rates of a micro-organism in a desired
PT environment, by introducing 2 reporter genes encoding luminescent and
PT fluorescent products and detecting luminescent fluorescence.
XX
PS Disclosure; Page 27; 32pp; English.
XX
XX The specification describes a method for assessing the growth rate and
CC death rate of a micro-organism within a predetermined time period in a
CC desired environment. The method comprises introducing at least two
CC reporter genes encoding luminescent and/or fluorescent products into the
CC micro-organism, incubating the micro-organism within the desired
CC environment, and detecting luminescence and/or fluorescence after a
CC predetermined time period. Use of two different markers within a micro-
CC organism enables the differentiation between growth and death rates. The
CC method is used to assess the growth rate and death rate of a micro-
CC organism within a predetermined time period in a desired environment. The
CC present sequence represents a green fluorescent protein (GFP), and is
CC encoded by a plasmid which encodes luminescent and fluorescent proteins,
CC and is used in the method of the invention
XX
SQ Sequence 239 AA;

Query Match 98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.3e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTILKFICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLTILKFICTTGKLPVWPWT 60
QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

RESULT 12
AAG66198
ID AAG66198 standard; protein; 239 AA.
XX
AC AAG66198;
XX
DT 17-JUN-2002 (first entry)
XX
XX A. victoria green fluorescent protein (EGFP).
DE
XX Cyan-green fluorescent protein; fluorescence; recombinant; GFP;
KW green fluorescent protein; EGFP.
XX
OS Aequorea victoria.
XX
PN JP2002045189-A.
XX
PD 12-FEB-2002.
XX

```

```
PF 04-AUG-2000; 2000JP-00237165.
XX
XX
XX 04-AUG-2000; 2000JP-00237165.
XX
XX (RIKA ) RIKAGAKU KENKYUSHO.
XX
XX WPI; 2002-299190/34.
XX
XX DR N-PSDB; ABL40628.
XX
XX
XX A gene encoding cyan-green fluorescent protein.
XX
XX Example; Page 14; 20pp; Japanese.
XX
XX The invention relates to a gene encoding proteins having cyan-green
CC fluorescence characteristic and having a function of showing stable
CC fluorescence characteristic in acid region. A method for the preparation
CC of a cyan-green fluorescent protein is provided which involves a
CC transformant transformed by a recombinant vector comprising the gene,
CC where the transformant is cultured and the protein is collected from the
CC culture. The present sequence represents the A. victoria green
CC fluorescent protein (EGFP)
XX
XX Sequence 239 AA;
SQ
Query Match 98.8%; Score 1258; DB 5; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.3e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
QY 61 LVTVLVSGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
RESULT 13
ABG94444
ID ABG94444 standard; protein; 239 AA.
XX
XX AC ABG94444;
XX
XX DT 27-NOV-2002 (first entry)
XX
XX DE Protease biosensor signal sequence #6.
XX
XX KW Detection; classification; identification; toxin detection; protease;
XX ADP-ribosylating toxin; cytotoxic phospholipase; exfoliative toxin;
XX toxic threat agent.
XX
XX OS Synthetic.
XX
XX PN US6416959-B1.
XX
XX PD 09-JUL-2002.
XX
XX PF 25-FEB-2000; 2000US-00513783.
XX
XX PR 27-FEB-1997; 97US-00810983.
XX PR 27-FEB-1998; 98US-00031271.
XX PR 26-FEB-1999; 99US-0122152P.
XX PR 08-MAR-1999; 99US-0123399P.
XX PR 12-JUL-1999; 99US-00352171.
XX PR 31-AUG-1999; 99US-0151797P.
PR 17-SEP-1999; 99US-00398965.
PR 29-OCT-1999; 99US-00430656.
PR 01-DEC-1999; 99US-0168408P.
XX
XX (GIUL/) GIULIANO K.
XX (KAPU/) KAPUR R.
XX
XX Giuliano K, Kapur R;
XX
XX WPI; 2002-634730/68.
XX DR N-PSDB; ABS71491.
XX
XX Automated cell-based toxin detection, classification, and/or
XX identification by treating cells involves use of three classes of
XX luminescent reporter molecules such as detectors, classifiers or
XX identifiers.
XX
XX Example 10; Fig 29A; 214pp; English.
XX
XX The invention describes methods of automated detection, classification
XX and identification comprising treating cells containing luminescent
XX reporter molecules (I) in array of locations with a test substance, where
XX (I) are detectors, classifiers or identifiers, imaging cells in each
XX location to obtain luminescent signals and converting optical information
XX into digital data to interpret presence of toxins in the test substance.
XX The method are useful for detection of toxins chosen from proteases, ADP-
XX ribosylating toxins, cytotoxic phospholipases, and exfoliative toxins.
XX Three classes of cell-based luminescent reporter molecules such as
XX detectors, classifiers and identifiers are described and serve as
XX reporters of toxic threat agents. The first two levels of
XX characterisation enable a rapid readout of toxin class without
XX sacrificing the ability to detect many new mutant toxins or dissect
XX several complex mixtures of known toxins. This is the amino acid sequence
XX of a protease biosensor related signal sequence used in the cell-based
XX screening system
XX
XX Sequence 239 AA;
SQ
Query Match 98.8%; Score 1258; DB 5; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.3e-121;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60
QY 61 LVTVLVSGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
RESULT 14
AAE14599
ID AAE14599 standard; protein; 239 AA.
XX
XX AC AAE14599;
XX
XX DT 31-MAY-2002 (first entry)
XX
XX DE Aequorea victoria enhanced green fluorescent protein.
XX KW Mutagenesis; enhanced green fluorescent protein; EGFP; mutant; mutein.
XX
XX OS Aequorea victoria.
XX Synthetic.
```



XX Key Location/Qualifiers  
 FT Misc-difference 1..3 /note= "Wild-type GFP Met-Ser are replaced with Met-Val-Ser"  
 FT FT  
 FT Misc-difference 65 /note= "GFP Phe64 is replaced by Leu"  
 FT FT  
 FT Misc-difference 66 /note= "GFP Ser65 is replaced by Thr"  
 FT FT  
 XX EPI178109-AL.  
 XX  
 XX 06-FEB-2002.  
 XX  
 XX 03-AUG-2001; 2001EP-00306650.  
 XX  
 XX 04-AUG-2000; 2000JP-00237166.  
 XX  
 XX (RIKE ) RIKEN KK.  
 XX  
 XX Miyawaki A, Sawano A;  
 XX WPI; 2002-208112/27.  
 XX N-PSDB; AAD27910.  
 XX  
 XX Method for mutagenesis, e.g. for introducing certain or random mutations at certain sites of the nucleotide sequence, comprises synthesizing a mutated strand and a complementary strand by use of megaprimers.  
 XX  
 XX Example 1; Page 13-14; 31pp; English.  
 XX  
 XX The invention relates to a method for mutagenesis that comprises synthesizing a mutated strand and a complementary strand by use of megaprimers. The method basically comprises a DNA synthesis in which one or more primers that have a nucleotide sequence containing at least one mutation and a phosphorylated 5'-terminus are annealed to a template DNA and then subjected to an elongation reaction using a thermostable high-fidelity DNA polymerase, after which the phosphorylated 5'-terminus and the elongated terminus are ligated by means of a thermostable DNA ligase to synthesise a circular DNA containing the primers; a digestion in which the step of DNA synthesis is repeated several times to amplify the DNA containing the primers and then, at least DNAs other than the amplified circular DNA are digested into several fragments; and a double-stranded DNA synthesis in which, with the several fragments, and a double-stranded of digestion as megaprimers, the megaprimers are annealed to the circular DNA synthesised above, followed by an elongation reaction performed using the thermostable high-fidelity DNA polymerase. The method is useful for mutagenesis, particularly for introducing certain mutations at certain sites of a nucleotide sequence, or for introducing random mutations at certain sites of the nucleotide sequence. The present method is simple, speedy, economical and widely applicable. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used for mutagenesis in an exemplification of the invention. The EGFP is derived by altering the green fluorescent protein (GFP) sequence of Aequorea victoria  
 XX  
 XX Sequence 239 AA;  
 XX  
 XX Query Match 98.8%; Score 1258; DB 5; Length 239;  
 XX Best Local Similarity 98.7%; Pred. No. 1.3e-121;  
 XX Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MVSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60  
 DB 1 MVSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60  
 QY 61 LVTLVSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDITL 120  
 DB 61 LVTTLTYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDITL 120  
 QY 121 VNRIELKGIDFKEDGNILGHLEYNVYNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 DB 121 VNRIELKGIDFKEDGNILGHLEYNVYNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 QY 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239  
 DB 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

QY 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239  
 DB 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239  
 RESULT 15  
 AAE34958  
 ID AAE34958 standard; protein; 239 AA.  
 XX  
 AC AAE34958;  
 XX  
 XX 28-MAY-2003 (first entry)  
 XX  
 DE Aequorea victoria enhanced green fluorescent protein (EGFP).  
 XX  
 KW Phosphorylation indicator; fluorescent protein; detection; phosphatase;  
 KW Kinase; enhanced green fluorescent protein; EGFP.  
 XX  
 OS Aequorea victoria.  
 XX  
 XX WO200295058-A2.  
 XX  
 XX 28-NOV-2002.  
 XX  
 XX 24-MAY-2002; 2002WO-US016955.  
 XX  
 XX 24-MAY-2001; 2001US-00865291.  
 XX  
 XX (REGC ) UNIV CALIFORNIA.  
 XX  
 XX Tsien RV, Ting AV, Zhang J;  
 XX WPI; 2003-148474/14.  
 XX N-PSDB; AAD53428.  
 XX  
 XX Novel chimeric phosphorylation indicators, useful for detecting kinase/phosphatase in samples, has donor molecule, phosphorylatable domain, phosphoaminoacid binding domain, and acceptor molecule, in operative linkage.  
 XX  
 XX Disclosure; Col 56-57; 38pp; English.  
 XX  
 XX The present invention relates to chimeric phosphorylation indicators comprising a phosphorylation polypeptide and a fluorescent protein or operative linkage, a donor molecule, a phosphorylatable domain, a phosphoaminoacid binding domain (PABD) and an acceptor molecule. The phosphorylation indicators of the invention are useful for detecting kinases or phosphatases in a biological sample. They are also useful in high throughput analysis e.g. for detecting a kinase inhibitor or phosphatase inhibitor. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used in the invention  
 XX  
 XX Sequence 239 AA;  
 XX  
 XX Query Match 98.8%; Score 1258; DB 6; Length 239;  
 XX Best Local Similarity 98.7%; Pred. No. 1.3e-121;  
 XX Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MVSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60  
 DB 1 MVSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60  
 QY 61 LVTLVSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDITL 120  
 DB 61 LVTTLTYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDITL 120  
 QY 121 VNRIELKGIDFKEDGNILGHLEYNVYNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 DB 121 VNRIELKGIDFKEDGNILGHLEYNVYNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 QY 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239  
 DB 181 DHYQONTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

Db 181 DHYQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMYLLEFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 15:54:09  
Job time : 47.1111 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:55 ; Search time 12.7778 Seconds  
(without alignments)  
965.630 Million cell updates/sec

Title: US-09-887-784-64v

Perfect score: 1273

Sequence: 1 MYSKGEELFTGVVPIVLVDL.....VLLGFVTRAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA:\*

- 1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/prodata/2/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/prodata/2/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/prodata/2/iaa/PCITUS\_COMB.pep.\*
- 6: /cgn2\_6/prodata/2/iaa/backfilesi.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1258	98.8	239	3	US-09-172-063-3 Sequence 3, Appli
2	1258	98.8	239	4	US-09-513-783A-46 Sequence 46, Appli
3	1258	98.8	239	4	US-09-316-919-4 Sequence 4, Appli
4	1258	98.8	239	4	US-09-602-641-3 Sequence 3, Appli
5	1258	98.8	239	4	US-09-920-922-2 Sequence 2, Appli
6	1258	98.8	281	3	US-09-062-102-1 Sequence 1, Appli
7	1258	98.8	281	4	US-09-364-946-1 Sequence 1, Appli
8	1258	98.8	294	4	US-09-513-783A-2 Sequence 2, Appli
9	1258	98.8	323	3	US-09-172-063-21 Sequence 21, Appli
10	1258	98.8	323	4	US-09-602-641-21 Sequence 21, Appli
11	1258	98.8	364	3	US-09-085-305-6 Sequence 6, Appli
12	1258	98.8	379	4	US-09-417-197-129 Sequence 129, App
13	1258	98.8	434	4	US-09-800-170-48 Sequence 48, Appli
14	1258	98.8	442	4	US-09-417-197-127 Sequence 127, App
15	1258	98.8	459	4	US-09-513-783A-170 Sequence 170, App
16	1258	98.8	544	4	US-09-417-197-113 Sequence 113, App
17	1258	98.8	544	4	US-09-417-197-115 Sequence 115, App
18	1258	98.8	604	4	US-09-417-197-59 Sequence 59, Appli
19	1258	98.8	605	4	US-09-417-197-41 Sequence 41, Appli
20	1258	98.8	606	4	US-09-417-197-65 Sequence 65, Appli
21	1258	98.8	607	4	US-09-417-197-47 Sequence 47, Appli
22	1258	98.8	630	4	US-09-417-197-63 Sequence 63, Appli
23	1258	98.8	631	4	US-09-417-197-39 Sequence 39, Appli
24	1258	98.8	633	4	US-09-417-197-45 Sequence 45, Appli
25	1258	98.8	635	4	US-09-417-197-125 Sequence 125, App
26	1258	98.8	642	2	US-08-818-253-2 Sequence 2, Appli
27	1258	98.8	642	2	US-08-818-253-6 Sequence 6, Appli

ALIGNMENTS

RESULT 1  
US-09-172-063-3  
; Sequence 3, Application US/09172063  
; Patent No. 6150176  
; GENERAL INFORMATION:  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Miyawaki, Atsushi  
; APPLICANT: Llopis, Juan  
; APPLICANT: Wachter, Rebekka M.  
; APPLICANT: Remington, S. James  
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
; FILE REFERENCE: 07257/071001  
; CURRENT APPLICATION NUMBER: US/09/172,063  
; CURRENT FILING DATE: 1998-10-13  
; EARLIER APPLICATION NUMBER: 09/094,359  
; EARLIER FILING DATE: 1998-06-09  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: EGFP  
US-09-172-063-3

Query Match 98.8%; Score 1258; DB 3; Length 239;  
Best Local Similarity 98.7%; Pred. No. 1.3e-126;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY	1	MYSKGEELFTGVVPIVLVDLGDVNGHKFSVSGEGDATYVGLTLKFICTTGKLPVWPWT	60
Db	1	MYSKGEELFTGVVPIVLVDLGDVNGHKFSVSGEGDATYVGLTLKFICTTGKLPVWPWT	60
QY	61	LVTVLTYGVQCFSRYPDHNKQHDFFKDSAMPEGVQERTIFFKDDGNGYKTRAEVKFEGDTL	120
Db	61	LVTTLTYGVQCFSRYPDHNKQHDFFKDSAMPEGVQERTIFFKDDGNGYKTRAEVKFEGDTL	120
QY	121	VNRIELKIDDFKEDGNILGHKLEYNNYSHNVIMADKQKNGIKVNFKIRHNIEDGSVOLA	180
Db	121	VNRIELKIDDFKEDGNILGHKLEYNNYSHNVIMADKQKNGIKVNFKIRHNIEDGSVOLA	180
QY	181	DHYQONTPTGDGPVLLPDNNHYLSTOSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK	239
Db	181	DHYQONTPTGDGPVLLPDNNHYLSTOSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK	239

```
RESULT 2
US-09-513-783A-46
; Sequence 46, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGFP
US-09-513-783A-46

Query Match          98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.3e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVSSEGGEGDATYKGLTLKFKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVSSEGGEGDATYKGLTLKFKICTTGKLPVPWPT 60
QY 51 LVTLSYGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 51 LVTTLTYGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 131 DHYQOQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLVLFVTAAGITLGMDELYK 239
DB 131 DHYQOQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLVLFVTAAGITLGMDELYK 239

RESULT 3
US-09-316-919-4
; Sequence 4, Application US/09316919
; Patent No. 6469154
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: 07257/073001
; CURRENT APPLICATION NUMBER: US/09/316,919
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-316-919-4

Query Match          98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.3e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVSSEGGEGDATYKGLTLKFKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVSSEGGEGDATYKGLTLKFKICTTGKLPVPWPT 60
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 181 DHYQOQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLVLFVTAAGITLGMDELYK 239
DB 181 DHYQOQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLVLFVTAAGITLGMDELYK 239

RESULT 4
US-09-602-641-3
; Sequence 3, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-602-641-3

Query Match          98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.3e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVSSEGGEGDATYKGLTLKFKICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVSSEGGEGDATYKGLTLKFKICTTGKLPVPWPT 60
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180
QY 181 DHYQOQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLVLFVTAAGITLGMDELYK 239
DB 181 DHYQOQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLVLFVTAAGITLGMDELYK 239

RESULT 5
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. 6673610
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; US-09-920-922-2

Query Match          98.8%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.3e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
    |||||
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
    |||||

QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTL 120
    |||||
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTL 120
    |||||

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
    |||||
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
    |||||

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
    |||||
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
    |||||

RESULT 6
US-09-062-102-1
; Sequence 1, Application US/09062102
; Patent No. 6130313
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangqiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100
; CURRENT APPLICATION NUMBER: US/09/062,102
; CURRENT FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: US 60/060,855
; EARLIER FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6130313
US-09-062-102-1

Query Match          98.8%; Score 1258; DB 3; Length 281;
Best Local Similarity 98.7%; Pred. No. 1.7e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
    |||||
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
    |||||

QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTL 120
    |||||
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTL 120
    |||||

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
    |||||
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
    |||||

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
    |||||
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
    |||||

RESULT 7
US-09-364-946-1
; Sequence 1, Application US/09364946
; Patent No. 6306600
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangqiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100CJP/D2
; CURRENT APPLICATION NUMBER: US/09/364,946
; CURRENT FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: US 09/191,233
; EARLIER FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6306600
US-09-364-946-1

Query Match          98.8%; Score 1258; DB 4; Length 281;
Best Local Similarity 98.7%; Pred. No. 1.7e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
    |||||
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
    |||||

QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTL 120
    |||||
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTL 120
    |||||

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
    |||||
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
    |||||

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
    |||||
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
    |||||

RESULT 8
US-09-513-783A-2
; Sequence 2, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GFP-DEVD-Annexin II construct
US-09-513-783A-2

Query Match          98.8%; Score 1258; DB 4; Length 294;
Best Local Similarity 98.7%; Pred. No. 1.8e-126;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
    |||||
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
    |||||
```

Db 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTAKFICTTGKLPVWPWT 60  
QY 61 LVTVLGYVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
Db 121 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 9

US-09-172-063-21  
; Sequence 21, Application US/09172063  
; Patent No. 6150176  
; GENERAL INFORMATION:  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Miyawaki, Atsushi  
; APPLICANT: Llopis, Juan  
; APPLICANT: Wachter, Rebekka M.  
; APPLICANT: Remington, S. James  
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
; FILE REFERENCE: 07257/071001  
; CURRENT APPLICATION NUMBER: US/09/172,063  
; CURRENT FILING DATE: 1998-10-13  
; EARLIER APPLICATION NUMBER: 09/094,359  
; EARLIER FILING DATE: 1998-06-09  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 21  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: GT-EGFP  
US-09-172-063-21

Query Match 98.8%; Score 1258; DB 3; Length 323;  
Best Local Similarity 98.7%; Pred. No. 2.1e-126;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTAKFICTTGKLPVWPWT 60  
Db 85 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTAKFICTTGKLPVWPWT 144  
QY 61 LVTVLGYVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 145 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 204  
QY 121 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
Db 205 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 264  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
Db 265 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 323

RESULT 10

US-09-602-641-21  
; Sequence 21, Application US/09602641  
; Patent No. 6608189  
; GENERAL INFORMATION:  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Miyawaki, Atsushi  
; APPLICANT: Llopis, Juan

; APPLICANT: Wachter, Rebekka M.  
; APPLICANT: Remington, S. James  
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
; FILE REFERENCE: 07257/071001  
; CURRENT APPLICATION NUMBER: US/09/602,641  
; CURRENT FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 09/172,063  
; PRIOR FILING DATE: 1998-10-13  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 21  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: GT-EGFP  
US-09-602-641-21

Query Match 98.8%; Score 1258; DB 4; Length 323;  
Best Local Similarity 98.7%; Pred. No. 2.1e-126;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTAKFICTTGKLPVWPWT 60  
Db 85 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTAKFICTTGKLPVWPWT 144  
QY 61 LVTVLGYVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 145 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 204  
QY 121 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
Db 205 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 264  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
Db 265 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 323

RESULT 11

US-09-085-305-6  
; Sequence 6, Application US/09085305  
; Patent No. 6191269  
; GENERAL INFORMATION:  
; APPLICANT: Pollock, Allan  
; APPLICANT: Lovett, David H.  
; APPLICANT: Turck, Johanna  
; TITLE OF INVENTION: Selective Induction of Apoptosis in  
; TITLE OF INVENTION: Malignant Cancer Cells by Delivery of N-Terminal  
; TITLE OF INVENTION: Interleukin-1-Alpha Pro-Piece Polypeptide  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Bozicevic & Reed, LLP  
; STREET: 285 Hamilton Ave, Suite 200  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/085,305  
; FILING DATE: 29-MAY-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Francis, Carol L  
REGISTRATION NUMBER: 36,513  
REFERENCE/DOCKET NUMBER: 6510/102US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-327-3400  
TELEFAX: 650-327-3231  
TELEX:  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 364 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-085-305-6

Query Match 98.8%; Score 1258; DB 3; Length 364;  
Best Local Similarity 98.7%; Pred. No. 2.5e-126;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYVKLTAKFICTTGKLPVPWPT 60  
DB 126 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYVKLTAKFICTTGKLPVPWPT 185

QY 61 LVTLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120  
DB 186 LVTTLTYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 245

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
DB 246 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 305

QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB 306 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 364

RESULT 12  
US-09-417-197-129  
Sequence 129, Application US/09417197  
Patent No. 6518021  
GENERAL INFORMATION:  
APPLICANT: Ole THASTRUP, et al.  
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An  
FILE OF INVENTION: On A Cellular Response  
FILE REFERENCE: 3759-0110P  
CURRENT APPLICATION NUMBER: US/09/417,197  
CURRENT FILING DATE: 1999-10-07  
NUMBER OF SEQ ID NOS: 143  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 129  
LENGTH: 379  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: actin-binding-domain-EGFP fusion  
US-09-417-197-129

Query Match 98.8%; Score 1258; DB 4; Length 379;  
Best Local Similarity 98.7%; Pred. No. 2.7e-126;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYVKLTAKFICTTGKLPVPWPT 60  
DB 141 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYVKLTAKFICTTGKLPVPWPT 200

QY 61 LVTLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120  
DB 201 LVTTLTYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 260

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
DB 261 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 320

QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB 321 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 379

RESULT 13  
US-09-800-170-48  
Sequence 48, Application US/09800170  
Patent No. 6481667  
GENERAL INFORMATION:  
APPLICANT: Kinsella, Todd  
TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES  
FILE REFERENCE: A-68614-1/DJB/RMS/RMK  
CURRENT APPLICATION NUMBER: US/09/800,170  
CURRENT FILING DATE: 2001-08-28  
PRIOR APPLICATION NUMBER: US 60/187,130  
PRIOR FILING DATE: 2000-03-06  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 48  
LENGTH: 434  
TYPE: PRT  
ORGANISM: Synechocystis PCC6803  
US-09-800-170-48

Query Match 98.8%; Score 1258; DB 4; Length 434;  
Best Local Similarity 98.7%; Pred. No. 3.3e-126;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYVKLTAKFICTTGKLPVPWPT 60  
DB 196 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYVKLTAKFICTTGKLPVPWPT 255

QY 61 LVTLSYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120  
DB 256 LVTTLTYGVQCFSRYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 315

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
DB 316 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGVSQVLA 375

QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB 376 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 434

RESULT 14  
US-09-417-197-127  
Sequence 127, Application US/09417197  
Patent No. 6518021  
GENERAL INFORMATION:  
APPLICANT: Ole THASTRUP, et al.  
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An  
FILE OF INVENTION: On A Cellular Response  
FILE REFERENCE: 3759-0110P  
CURRENT APPLICATION NUMBER: US/09/417,197  
CURRENT FILING DATE: 1999-10-07  
NUMBER OF SEQ ID NOS: 143  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 127  
LENGTH: 442  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: EGFP-RhoA fusion  
US-09-417-197-127

Query Match 98.8%; Score 1258; DB 4; Length 442;  
Best Local Similarity 98.7%; Pred. No. 3.4e-126;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYVKLTAKFICTTGKLPVPWPT 60

```
Db 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60
Qy 61 LVTLSYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
Db 61 LVTTLTYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120
Qy 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIWADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIWADKQNGIKVNFKIRHNIEDGSVQLA 180
Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239
```

RESULT 15  
US-09-513-783A-170  
; Sequence 170, Application US/09513783A  
; Patent No. 6416959  
; GENERAL INFORMATION:  
; APPLICANT: Giuliano, Kenneth A.  
; APPLICANT: Kapur, Ravi  
; TITLE OF INVENTION: A System for Cell Based Screening  
; FILE REFERENCE: 97-022-L1  
; CURRENT APPLICATION NUMBER: US/09/513,783A  
; CURRENT FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 170  
; LENGTH: 459  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP27  
US-09-513-783A-170

Query Match 98.8%; Score 1258; DB 4; Length 459;  
Best Local Similarity 98.7%; Pred. No. 3.5e-126;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Caps 0;  
Qy 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60  
Db 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60  
Qy 61 LVTLSYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120  
Db 61 LVTTLTYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNGYKTRAEVKFEGDTL 120  
Qy 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIWADKQNGIKVNFKIRHNIEDGSVQLA 180  
Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIWADKQNGIKVNFKIRHNIEDGSVQLA 180  
Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:04:01  
Job time : 12.7778 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:54:20 ; Search time 34.7778 Seconds  
(without alignments)  
1940.117 Million cell updates/sec

Title: US-09-887-784-64v

Perfect score: 1273

Sequence: 1 MVSKEELFTGVVPIVLVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA:\*
- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
  - 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
  - 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
  - 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
  - 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
  - 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
  - 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
  - 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
  - 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
  - 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
  - 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
  - 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1269	99.7	239	9	US-09-887-784-4
2	1269	99.7	239	12	US-10-296-953-4
3	1269	99.7	363	14	US-10-270-223-6
4	1269	99.7	893	14	US-10-257-909A-30
5	1269	99.7	1132	14	US-10-257-909A-32
6	1261	99.1	239	9	US-09-887-784-2
7	1261	99.1	239	12	US-10-296-953-2
8	1258	98.8	239	9	US-09-920-922-2
9	1258	98.8	239	9	US-09-999-745-4
10	1258	98.8	239	10	US-09-866-538-4
11	1258	98.8	239	10	US-09-797-496B-2
12	1258	98.8	239	10	US-09-794-308-4
13	1258	98.8	239	10	US-09-865-291-4
14	1258	98.8	239	12	US-10-457-982-3
15	1258	98.8	239	14	US-10-121-258-13

16	1258	98.8	239	14	US-10-221-461-7	Sequence 7, Appli
17	1258	98.8	239	14	US-10-100-957A-46	Sequence 46, Appli
18	1258	98.8	239	14	US-10-177-390-2	Sequence 2, Appli
19	1258	98.8	239	14	US-10-338-411-3	Sequence 3, Appli
20	1258	98.8	239	15	US-10-370-570-4	Sequence 4, Appli
21	1258	98.8	239	15	US-10-389-640-3	Sequence 3, Appli
22	1258	98.8	259	14	US-10-314-861-11	Sequence 11, Appli
23	1258	98.8	281	12	US-09-931-232-1	Sequence 1, Appli
24	1258	98.8	288	14	US-10-314-861-37	Sequence 37, Appli
25	1258	98.8	293	14	US-10-314-861-35	Sequence 35, Appli
26	1258	98.8	294	14	US-10-100-957A-2	Sequence 2, Appli
27	1258	98.8	295	14	US-10-314-861-39	Sequence 39, Appli
28	1258	98.8	299	14	US-10-314-861-33	Sequence 33, Appli
29	1258	98.8	305	14	US-10-314-861-31	Sequence 31, Appli
30	1258	98.8	308	14	US-10-033-717-35	Sequence 35, Appli
31	1258	98.8	311	14	US-10-314-861-29	Sequence 29, Appli
32	1258	98.8	320	14	US-10-338-411-11	Sequence 11, Appli
33	1258	98.8	320	15	US-10-389-640-11	Sequence 11, Appli
34	1258	98.8	323	12	US-10-457-982-21	Sequence 21, Appli
35	1258	98.8	323	14	US-10-338-411-7	Sequence 7, Appli
36	1258	98.8	323	14	US-10-338-411-13	Sequence 13, Appli
37	1258	98.8	323	15	US-10-389-640-7	Sequence 7, Appli
38	1258	98.8	323	15	US-10-389-640-13	Sequence 13, Appli
39	1258	98.8	324	14	US-10-314-861-16	Sequence 16, Appli
40	1258	98.8	345	14	US-10-338-411-5	Sequence 5, Appli
41	1258	98.8	345	15	US-10-389-640-5	Sequence 5, Appli
42	1258	98.8	346	14	US-10-338-411-9	Sequence 9, Appli
43	1258	98.8	346	15	US-10-389-640-9	Sequence 9, Appli
44	1258	98.8	359	14	US-10-033-717-33	Sequence 33, Appli
45	1258	98.8	359	14	US-10-033-717-34	Sequence 34, Appli

ALIGNMENTS

RESULT 1

US-09-887-784-4  
; Sequence 4, Application US/09887784  
; Patent No. US20020177189A1  
; GENERAL INFORMATION:  
; APPLICANT: BIORN, Sara et al  
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS  
; FILE REFERENCE: 3759-0115P  
; CURRENT APPLICATION NUMBER: US/09/887,784  
; CURRENT FILING DATE: 2001-06-19  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 4  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequoria Victoria  
US-09-887-784-4

Query Match 99.7%; Score 1269; DB 9; Length 239;  
Best Local Similarity 99.6%; Pred. No. 7.3e-124;  
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MVSKEELFTGVVPIVLVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGTGLPVPWPT	60
Db	1	MVSKEELFTGVVPIVLVELDGVNGHKFSVSGEGDATYVKLTAKFICTTGTGLPVPWPT	60
Qy	61	LVTVLISYGVQCFSRYPDHMKQHDFFPKSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL	120
Db	61	LVTVLISYGVQCFSRYPDHMKQHDFFPKSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL	120
Qy	121	VNRIELKIDFDKEDGNILGHKLEYNHNYSNHYIMADKQKNGIKVNFKIRHNIEDGVSQLA	180
Db	121	VNRIELKIDFDKEDGNILGHKLEYNHNYSNHYIMADKQKNGIKVNFKIRHNIEDGVSQLA	180
Qy	181	DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVLGFTVTAAGITLGMDELYK	239
Db	181	DHYQNTPTIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVLGFTVTAAGITLGMDELYK	239

```
RESULT 2
US-10-296-953-4
; Sequence 4, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-4

Query Match          99.7%; Score 1269; DB 12; Length 239;
Best Local Similarity 99.6%; Pred. No. 7.3e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYVKLTLPKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYVKLTLPKFICTTGKLPVPWPT 60
QY 51 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDYL 120
DB 51 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDYL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 131 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITIGMDLYK 239
DB 131 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITIGMDLYK 239

RESULT 3
US-10-270-223-6
; Sequence 5, Application US/10270223
; Publication No. US2003014363A1
; GENERAL INFORMATION:
; APPLICANT: BioImage A/S
; TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR COMPONENTS
; TITLE OF INVENTION: INTERACT LIVING CELLS, AND TO EXTRACT QUANTITATIVE INFORMATION RE
; FILE REFERENCE: 3759-0126P
; CURRENT APPLICATION NUMBER: US/10/270,223
; CURRENT FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Aequorea Victoria and Human
US-10-270-223-6

Query Match          99.7%; Score 1269; DB 14; Length 363;
Best Local Similarity 99.6%; Pred. No. 1.3e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYVKLTLPKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYVKLTLPKFICTTGKLPVPWPT 60
QY 51 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDYL 120
DB 51 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDYL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 131 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITIGMDLYK 239
DB 131 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITIGMDLYK 239

RESULT 4
US-10-257-909A-30
; Sequence 30, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 893
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-30

Query Match          99.7%; Score 1269; DB 14; Length 893;
Best Local Similarity 99.6%; Pred. No. 4.8e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYVKLTLPKFICTTGKLPVPWPT 60
DB 655 MVSKEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYVKLTLPKFICTTGKLPVPWPT 714
QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDYL 120
DB 715 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDYL 774
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 775 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 834
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITIGMDLYK 239
DB 835 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITIGMDLYK 893

RESULT 5
US-10-257-909A-32
; Sequence 32, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSEQ for Windows Version 3.0
```

```
; SEQ ID NO 32
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-32

Query Match          99.7%; Score 1269; DB 14; Length 1132;
Best Local Similarity 99.6%; Pred. No. 6 8e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLTKFICTTGKLPVPWPT 60
Db 894 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLTKFICTTGKLPVPWPT 953

Qy 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 954 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQVQERTIFFKDDGNYKTRAEVKFEGDTL 1013

Qy 121 VNRIELKGIDFDKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 1014 VNRIELKGIDFDKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 1073

Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
Db 1074 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 1132

RESULT 6
US-09-887-784-2
; Sequence 2, Application US/09887784
; Patent No. US2002017189A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea Victoria
US-09-887-784-2

Query Match          99.1%; Score 1261; DB 9; Length 239;
Best Local Similarity 99.2%; Pred. No. 5e-123;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLTKFICTTGKLPVPWPT 60
Db 1 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLTKFICTTGKLPVPWPT 60

Qy 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQVQERTIFFKDDGNYKTRAEVKFEGDTL 120

Qy 121 VNRIELKGIDFDKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFDKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 7
US-10-296-953-2
; Sequence 2, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match          99.1%; Score 1261; DB 12; Length 239;
Best Local Similarity 99.2%; Pred. No. 5e-123;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLTKFICTTGKLPVPWPT 60
Db 1 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLTKFICTTGKLPVPWPT 60

Qy 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQVQERTIFFKDDGNYKTRAEVKFEGDTL 120

Qy 121 VNRIELKGIDFDKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFDKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 8
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. US20020083488A1
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match          98.8%; Score 1258; DB 9; Length 239;
Best Local Similarity 98.7%; Pred. No. 1e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLTKFICTTGKLPVPWPT 60
Db 1 MWSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLTKFICTTGKLPVPWPT 60
```

QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQKNGIKVNFKIRHNIEDGSGVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQKNGIKVNFKIRHNIEDGSGVQLA 180  
QY 181 DHYQOQNTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQOQNTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLLGFVTAAGITLGMDELYK 239

RESULT 9  
US-09-999-745-4  
; Sequence 4, Application US/09999745  
; Patent No. US2002015120A1  
; GENERAL INFORMATION:  
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Baird, Geoffrey  
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS  
; FILE REFERENCE: REGEN1470-1  
; CURRENT APPLICATION NUMBER: US/09/999,745  
; CURRENT FILING DATE: 2001-10-23  
; PRIOR APPLICATION NUMBER: 09/316,920  
; PRIOR FILING DATE: 1999-05-21  
; NUMBER OF SEQ ID NOS: 67  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
US-09-999-745-4

Query Match 98.8%; Score 1258; DB 9; Length 239;  
Best Local Similarity 98.7%; Pred. No. 1e-122;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60  
QY 51 LVTLSYGVCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 51 LVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQKNGIKVNFKIRHNIEDGSGVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQKNGIKVNFKIRHNIEDGSGVQLA 180  
QY 181 DHYQOQNTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQOQNTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLLGFVTAAGITLGMDELYK 239

RESULT 10  
US-09-866-538-4  
; Sequence 4, Application US/09866538  
; Publication No. US20030032088A1  
; GENERAL INFORMATION:  
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA  
; APPLICANT: Campbell, Robert  
; APPLICANT: Tsien, Roger  
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS  
; FILE REFERENCE: REGEN1530-2  
; CURRENT APPLICATION NUMBER: US/09/866,538  
; CURRENT FILING DATE: 2001-05-24  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 239  
; TYPE: PRT

; ORGANISM: Aequorea victoria  
US-09-866-538-4  
Query Match 98.8%; Score 1258; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 1e-122;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60  
QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQKNGIKVNFKIRHNIEDGSGVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQKNGIKVNFKIRHNIEDGSGVQLA 180  
QY 181 DHYQOQNTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQOQNTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLLGFVTAAGITLGMDELYK 239

RESULT 11  
US-09-797-496B-2  
; Sequence 2, Application US/09797496B  
; Publication No. US20030049597A1  
; GENERAL INFORMATION:  
; APPLICANT: Simon, Sanford M.  
; APPLICANT: Chen, Yu  
; TITLE OF INVENTION: Chimeric Fluorescent Enzymes and Uses Thereof  
; FILE REFERENCE: 600-1-267  
; CURRENT APPLICATION NUMBER: US/09/797,496B  
; CURRENT FILING DATE: 2002-05-24  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Aequorea victoria green fluorescent protein modified as described  
; OTHER INFORMATION: in specification  
US-09-797-496B-2

Query Match 98.8%; Score 1258; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 1e-122;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60  
QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQKNGIKVNFKIRHNIEDGSGVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQKNGIKVNFKIRHNIEDGSGVQLA 180  
QY 181 DHYQOQNTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQOQNTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLLGFVTAAGITLGMDELYK 239

RESULT 12  
US-09-794-308-4  
; Sequence 4, Application US/09794308  
; Publication No. US20030170911A1  
; GENERAL INFORMATION:  
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA

APPLICANT: TSJEN, Roger  
APPLICANT: ZACHARIAS, David  
APPLICANT: BAIRD, Geoffrey  
TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS  
FILE REFERENCE: REGEN1530  
CURRENT APPLICATION NUMBER: US/09/794,308  
CURRENT FILING DATE: 2001-02-26  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
US-09-794-308-4

Query Match 98.8%; Score 1258; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 1e-122; Indels 0; Gaps 0;  
Matches 236; Conservative 1; Mismatches 2;

QY 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDGYGKLTTLKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDGYGKLTTLKFICTTGKLPVPWPT 60  
QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGVPLLPDNNHLYSTQSALSKDPNEKRDHMLVGLFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGVPLLPDNNHLYSTQSALSKDPNEKRDHMLVGLFVTAAGITLGMDELYK 239

RESULT 13  
US-09-865-291-4  
Sequence 4, Application US/09865291  
Publication No. US20030186229A1  
GENERAL INFORMATION:  
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA  
APPLICANT: TSJEN, Roger  
APPLICANT: TING, Alice  
APPLICANT: ZHANG, Jin  
TITLE OF INVENTION: EMISSION RATIONETRIC INDICATORS OF PHOSPHORYLATION  
FILE REFERENCE: REGEN1550  
CURRENT APPLICATION NUMBER: US/09/865,291  
CURRENT FILING DATE: 2001-05-24  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
US-09-865-291-4

Query Match 98.8%; Score 1258; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 1e-122;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDGYGKLTTLKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDGYGKLTTLKFICTTGKLPVPWPT 60  
QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGVPLLPDNNHLYSTQSALSKDPNEKRDHMLVGLFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGVPLLPDNNHLYSTQSALSKDPNEKRDHMLVGLFVTAAGITLGMDELYK 239

RESULT 14  
US-10-457-982-3  
Sequence 3, Application US/10457982  
Publication No. US20030212265A1  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger Y.  
APPLICANT: Miyawaki, Atsushi  
APPLICANT: Llopis, Juan  
APPLICANT: Wächter, Rebekka M.  
APPLICANT: Remington, S. James  
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
MEASURING THE PH OF A BIOLOGICAL SAMPLE  
FILE REFERENCE: 07257/071001  
CURRENT APPLICATION NUMBER: US/10/457,982  
CURRENT FILING DATE: 2003-06-09  
PRIOR APPLICATION NUMBER: US/09/602,641  
PRIOR FILING DATE: 2000-06-22  
PRIOR APPLICATION NUMBER: 09/172,063  
PRIOR FILING DATE: 1998-10-13  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (0)...(0)  
OTHER INFORMATION: EGFP  
US-10-457-982-3

Query Match 98.8%; Score 1258; DB 12; Length 239;  
Best Local Similarity 98.7%; Pred. No. 1e-122; Indels 0; Gaps 0;  
Matches 236; Conservative 1; Mismatches 2;

QY 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDGYGKLTTLKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDGYGKLTTLKFICTTGKLPVPWPT 60  
QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGVPLLPDNNHLYSTQSALSKDPNEKRDHMLVGLFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGVPLLPDNNHLYSTQSALSKDPNEKRDHMLVGLFVTAAGITLGMDELYK 239

RESULT 15  
US-10-121-258-13  
Sequence 13, Application US/10121258  
Publication No. US20030059835A1  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger  
APPLICANT: Campbell, Robert  
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT  
PROTEIN VARIANTS AND METHODS FOR MAKING SAME  
FILE REFERENCE: UC083.1CP2CP1  
CURRENT APPLICATION NUMBER: US/10/121,258  
CURRENT FILING DATE: 2002-04-10  
PRIOR APPLICATION NUMBER: 09/794,308  
PRIOR FILING DATE: 2001-02-26  
PRIOR APPLICATION NUMBER: 09/866,538  
PRIOR FILING DATE: 2001-05-24

```
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Enhanced Green Fluorescent Protein (EGFP)
US-10-121-258-13

Query Match      98.8%; Score 1258; DB 14; Length 239;
Best Local Similarity 98.7%; Pred. No. 1e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 MVSKEELFTGVVPILVELDGDVNGHKFSVSSEGEEDATYKLTTLKFICTTGKLPVPWPT 60
Db      1 MVSKEELFTGVVPILVELDGDVNGHKFSVSSEGEEDATYKLTTLKFICTTGKLPVPWPT 60

QY      61 LVTLSVGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db      61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY      121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db      121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY      181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLGFVTAAGITLGMDELYK 239
Db      181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLGFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:09:27
Job time : 35.7778 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:00 ; Search time 10.3333 Seconds  
(without alignments)  
2224.817 Million cell updates/sec

Title: US-09-887-784-64V  
Perfect score: 1273  
Sequence: 1 MVSKEELFTGVVPILVELD.....VLLGFVTRAGITLGMDELYK 239  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1236	97.1	238	1 JQ1514	green-fluorescent
2	1105	8.2	785	2 H72228	hypothetical prote
3	93.5	7.3	861	2 H64102	leucine-trRNA ligase
4	91.5	7.2	887	2 E82590	leucyl-trRNA synthetase
5	89.5	7.0	860	2 AC0582	DNA topoisomerase
6	88.5	7.0	655	2 D83917	iron-sulfur cofactor
7	87.5	6.9	370	2 E70390	hypothetical prote
8	87.5	6.9	2573	2 D71614	DNA-directed DNA p
9	86.5	6.8	788	1 JCVLHH	protective surface
10	86.5	6.8	797	2 JDC4078	protective surface
11	86.5	6.8	808	2 F64102	tRNA (uracil-5)-m
12	86	6.8	357	2 G81355	photosystem II chl
13	86	6.8	461	2 T06936	inter-alpha-trypsi
14	85.5	6.7	889	2 JC5576	cellulase (EC 3.2.
15	85.5	6.7	941	2 S29043	hypothetical prote
16	85	6.7	281	2 AD2052	hypothetical prote
17	85	6.7	336	2 C64468	synaptogamin o-p65
18	85	6.7	439	2 JH0414	DNA-binding protei
19	85	6.7	632	2 T06586	oligodendrocyte
20	84.5	6.6	613	2 A95552	DNA-directed DNA p
21	84.5	6.6	836	1 JCVLDD	thioredoxin reduct
22	84	6.6	285	2 C97279	dihydrofolamide d
23	84	6.6	578	1 I40794	hypothetical prote
24	84	6.6	682	2 D90946	hypothetical prote
25	84	6.6	682	2 H85794	hypothetical prote
26	84	6.6	682	2 A47198	carboxy-terminal p
27	83.5	6.6	425	2 C97354	hypothetical prote
28	83.5	6.6	1134	2 A60234	IgA Fc receptor pr
29	83.5	6.6	1164	1 FC50AG	IgA Fc receptor pr

RESULT 1  
JQ1514

Green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)  
C:Species: Aequorea victoria  
C:Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 23-Mar-2001  
C:Accession: JQ1514; PQ0335; S48693; S51330; S51331  
R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.  
Gene 111, 229-233, 1992  
A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.  
A:Reference number: JQ1514; MUID:92175527; PMID:1347277  
A:Accession: JS0692  
A:Molecule type: DNA  
A:Residues: 1-107, 'S', 109-238 <PEAL>  
A:Cross-references: GB:M62654; NID:G155662; PIDN:AAA27722.1; PID:G155663  
A:Accession: JQ1514  
A:Molecule type: mRNA  
A:Residues: 1-99, 'P', 'L', 101-140, 'L', 142-218, 'V', 220-238 <PRA2>  
A:Cross-references: GB:M62653; NID:G155660; PIDN:AAA27721.1; PID:G155661  
A:Accession: PQ0335

ALIGNMENTS

30 82.5 6.5 860 1 SYECL leucine-trRNA ligase  
31 82.5 6.5 860 2 H90713 leucine trRNA synth  
32 82.5 6.5 860 2 D85564 leucine trRNA synth  
33 82.5 6.5 2222 1 A36028 DNA-directed DNA p  
34 82 6.4 353 2 E84941 imidazoleglycerol-  
35 82 6.4 471 2 T27856 hypochlorite prote  
36 82 6.4 752 1 KXRTCL propeptin convert  
37 82 6.4 764 2 T25012 hypochlorite prote  
38 82 6.4 774 2 T39539 alpha-amylase homo  
39 82 6.4 865 2 T11852 lipoxigenase (EC 1  
40 82 6.4 874 2 JC4930 S-layer protein pr  
41 81.5 6.4 263 2 S53488 water-stress-induc  
42 81.5 6.4 740 2 G95153 neuraminidase, pro  
43 81 6.4 310 2 S68225 synergohymenotrophi  
44 81 6.4 355 2 H97144 ferrichrome-bindin  
45 81 6.4 529 2 B66815 hypothetical prote





A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A;Reference number: A59328  
A;Contents: annotation  
C;Genetics:  
A;Gene: XF2176  
C;Superfamily: leucine-tRNA ligase

Query Match 7.2%; Score 91.5; DB 2; Length 887;  
Best Local Similarity 22.2%; Pred. No. 14;  
Matches 44; Conservative 30; Mismatches 67; Indels 57; Gaps 10;

QY 50 TTGKLPVPMPTLVLSYGVCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDCGNY-- 107  
Db TNEQLPV-WVANPVMAYGTGAVMAVEGHQDQDEF--ANKYGLPIRQVIALKEPKNQDE 385

QY 108 -----KTRAEVKFEGDTLVNRIELKGDIDFKEDGNILGHKLEYNNSHNVI 153  
Db STWEPDWRDWAYDKTR---EFE---LINSAEFDGLDYQDAFEVLAERFE----- 429

QY 154 MADKQKNG-IKVNFKIRHNIEDGSVOLADHYQQNTPI-----GDGPVLLPDN 199  
Db 430 ---RQGRQRVNYRLR-----DWGVSQRQYWGCPPIVIYCTCGAVVPEDQLPVILPEN 482

QY 200 -HYLSTQSALSADPNRKR 216  
Db 483 VAFSGTGSPIKTDPEWRK 500

RESULT 5  
AC0582  
leucyl-tRNA synthetase [imported] - Salmonella enterica subsp. enterica serovar Typhi (S  
A;Species: Salmonella enterica subsp. enterica serovar Typhi  
A;Note: this species has also been called salmonella typhi  
C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C;Accession: AC0582  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
th, T.; Connor, A.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A;Reference number: AB0502; MUID:21534947; PMID:11677608  
A;Accession: AC0582  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-860 <PAR>  
A;Cross-references: GB:AL513382; PIDN:CAD05125.1; PID:g16501899; GSPDB:GN00176  
C;Genetics:  
A;Gene: STY0699  
C;Superfamily: leucine-tRNA ligase

Query Match 7.0%; Score 89.5; DB 2; Length 860;  
Best Local Similarity 23.3%; Pred. No. 20;  
Matches 42; Conservative 22; Mismatches 77; Indels 39; Gaps 7;

QY 50 TTGKLPVPMPTLVLSYGVCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDCGNYKT 109  
Db TGEIPIV-WAANPVMLEYGTGAVMAVFGH-QDQD-YEFASKYGLTIKPVILADGSEPD 370

QY 110 RAEVKEGDTLVNRIELKGDIDFKEDGNILGHKLEYNNSHNVIKQNGIKVNFKIR 169  
Db SEQALTEKGVLFNSGFDGLAFEAFAATADKL-----AEKVGGERKVNRYLR 418

QY 170 H-----NIEDGSVOLADHYQQNTPIGDGPVLLPDNHYL-STOSALSADP 212  
Db DWGVSQRQYWGAPIPWVLTEDGT-----LPTPDQLPVILPDWVMDGITSPIKADP 471

RESULT 6  
DB3917  
DNA topoisomerase IV subunit B BH2140 [imported] - Bacillus halodurans (strain C-125)  
C;Species: Bacillus halodurans

C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C;Accession: DB3917  
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000  
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A;Reference number: AB3650; MUID:20512582; PMID:11058132  
C;Accession: DB3917  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-655 <STO>  
A;Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BA05859.1; GSPDB:GN00  
A;Experimental source: strain C-125  
C;Genetics:  
A;Gene: BH2140  
C;Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B

Query Match 7.0%; Score 88.5; DB 2; Length 655;  
Best Local Similarity 21.9%; Pred. No. 17;  
Matches 53; Conservative 37; Mismatches 95; Indels 57; Gaps 9;

QY 22 DVNGHK---FSVSGEGEGDAT---YKLTLLKFI-----CTTGKLPVPWP 59  
Db 63 NVTIHKQSVSRVDEGRGMPGTGMHKLKGTPEVILTVLHAGGKFGGGYATSGLHCVGA 122

QY 60 TLVTLSYGVCFSRYPDHMKQHDFFKSAPEGYVQER-----TIFKDDG----- 105  
Db 123 SVNALSEWLIVEIKRDGWVQEPFENGKPKSTLEKKGKTRGTGTTIHKPDPTVFSTT 182

QY 106 --NYKTRAEVKFEGDTLVNRIELKGDIDFKEDGNILGHKLEYNNSHNVIYMAK----- 157  
Db 183 NFNVTLSRLREAAFLKGLKIELVDLRDQDTEKVFH-YEDGIGKAFVEYLNEDEKTLHPV 241

QY 158 ----QKNGIKVNFKIRHNIEDGSVOLADHYQQNTPIGDGPVLLPDNHYLSTOSALSADP 212  
Db 242 VFNNGESNGIEIEFAFQFN--DQYTNVLSFVNVTQDQ-----GTHELCAKTMTRAV 294

QY 213 NE 214  
Db 295 NE 296

RESULT 7  
E70390  
iron-sulfur cofactor synthesis protein nifs - Aquifex aeolicus  
N;Contains: L-cysteine sulfotransferase (EC 2.8.1.-)  
C;Species: Aquifex aeolicus  
C;Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 07-Dec-1999  
C;Accession: E70390  
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove  
V.  
Nature 392, 353-358, 1998  
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A;Reference number: A70300; MUID:98196666; PMID:9537320  
A;Accession: E70390  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-370 <AQF>  
A;Cross-references: GB:AE000720; NID:g2983529; PIDN:AAC07111.1; PID:g2983536; GB:AE00065  
C;Genetics:  
A;Gene: nifs1  
C;Superfamily: nitrogen fixation protein nifs  
C;Keywords: phosphoprotein; pyridoxal phosphate; sulfotransferase  
F;195/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted  
F;318/Active site: Cys (cysteine persulfide intermediate) #status predicted

Query Match 6.9%; Score 87.5; DB 2; Length 370;  
Best Local Similarity 25.4%; Pred. No. 9.8;  
Matches 49; Conservative 30; Mismatches 91; Indels 23; Gaps 8;

QY 4 KGEELFTGVV----PILVELD-----GDVNGHKF-SVSGEG-----EGDATYKLTLPFICT 50  
Db 164 KGVPLLTDVAQIKGPIELNISATFSCHKFHAIKSGFLYISDEANYEPLIVGGQE 223

```
QY 51 TGKLP-----VPMPTLVLSYGVQCFSRYPDHMKQ-HDFKSAPEGYVQERTIFFKOD 104
Db 224 NGRKSGTENVGILSLAKALEIIVNSFRYQEQKKURDUFENLLLEA-LPDAQIVGKDA 282
QY 105 GNYKTRAEV---RFGEDTLVNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNG 161
Db 283 ERSPSISVIMPRFFGAEIVNKLSEKGIYCGTSGACLSGEYENKMLKMGFSQEKALRM 342
QY 162 IKVNFKIRHNIED 174
Db 343 VRFSFGLLNKEEE 355

RESULT 8
D71614
hypothetical protein PFB0460c - malaria parasite (Plasmodium falciparum)
C/Species: Plasmodium falciparum
C/Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C/Accession: D71614
R/Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Percec, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;
Science 282, 1126-1132, 1998
A/Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A/Reference number: A71600; MUID:99021743; PMID:9804551
A/Accession: D71614
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-2573 <GAR>
A/Cross-references: GB:AE001396; GB:AE001362; NID:g3845188; PIDN:AACT71881.1; PID:g384519
A/Experimental source: clone 3D7
C/Genetics:
A/Gene: PFB0460c

Query Match 6.9%; Score 87.5; DB 2; Length 2573;
Best Local Similarity 26.2%; Pred. No. 1.2e+02;
Matches 34; Conservative 30; Mismatches 53; Indels 13; Gaps 5;

QY 94 VQERTIFFKD--DNGYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYN--YNH 149
Db 126 LKKTETLCKDIKSGNDPDEISLFDKDDWDDKELK--DFEKSLLKKNKEVNFYNNK 183
QY 150 NVYIMADKQKNGIKVNFKIRHNIEDGSGVQLADHYQNTPIGDGPVLLPNHLYLSTOSALS 209
Db 134 NLHIKENKKDEKKNKHNNNDNNM-----IYYKNI---DKTHYLDNNVVHILNDIN 236
QY 210 KDPNEKRDM 219
Db 237 TYLKRERDYM 246

RESULT 9
JDVLHH
DNA-directed DNA polymerase (EC 2.7.7.7) - heron hepatitis virus
C/Species: heron hepatitis virus, HHV9
A/Note: host: Ardea cinerea (Gray heron)
C/Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 11-Jun-1999
C/Accession: A30082
R/Sprenger, R.; Kaleta, E.F.; Will, H.
J. Virol. 62, 3832-3839, 1988
A/Title: Isolation and characterization of a hepatitis B virus endemic in herons.
A/Reference number: A93037; MUID:88333160; PMID:3418788
C/Accession: A30082
A/Molecule type: DNA
A/Residues: 1-788 <SPR>
A/Cross-references: GB:M22056; NID:g325452; PIDN:AAA45738.1; PID:g325454
C/Superfamily: hepatitis virus DNA-directed DNA polymerase
C/Keywords: DNA biosynthesis; nucleotidyltransferase

Query Match 6.8%; Score 86.5; DB 1; Length 788;
Best Local Similarity 19.7%; Pred. No. 32;
Matches 40; Conservative 32; Mismatches 66; Indels 65; Gaps 10;
```

```
QY 58 WPTLVLSYGVQCFSRYPDHMKQ-----DFFKSAPEGYVQERT-----IFFKDDGNYK 108
Db 139 WPKSISTPLVHSGVKPKYFEFQNHESLVNDYLNKLFEAGILYKRVSKHLVTFK--GPYP 196
QY 109 T-----RAEVKFGDTLVNRIELKGIDFKEDGNILGHKLEYNYNHNVYIM 154
Db 197 TWEQKHLVPQOHGAYSSKINDRQESRRRIITATSSRKNDSSRI-----FCAHN--- 245
QY 155 ADKQKNGIKVNFKIRHNIEDGSGVQLADHYQ-----QNTPIGDGPVLL--PNHYL 202
Db 246 -----NGRKISY---HSTRDGGHRLSGRTSDPTSRGALAGSDSTPGCGTAHPSTHHV 297
QY 203 STQ-----SALSKDPNEKR 216
Db 298 DRRRROKGGVQLQAISREFSETR 320

RESULT 10
JC4078
protective surface antigen D-15 precursor - Haemophilus influenzae (type b)
C/Species: Haemophilus influenzae
A/Variety: type b
C/Date: 30-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 29-Sep-1999
C/Accession: JC4078
R/Flack, F.S.; Loosmore, S.; Chong, P.; Thomas, W.R.
Gene 156, 97-99, 1995
A/Title: The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus infl
A/Reference number: JC4078; MUID:95255676; PMID:7737523
A/Accession: JC4078
A/Molecule type: DNA
A/Residues: 1-797 <FLA>
A/Cross-references: GB:U13961; NID:g537447; PIDN:AAA85645.1; PID:g537448
A/Experimental source: type b
C/Superfamily: protective surface antigen D-15
C/Keywords: surface antigen
F/1-19/Domain: signal sequence #status predicted <SIG>
F/20-797/Product: protective surface antigen D-15 #status predicted <MAT>

Query Match 6.8%; Score 86.5; DB 2; Length 797;
Best Local Similarity 21.9%; Pred. No. 32;
Matches 48; Conservative 30; Mismatches 78; Indels 63; Gaps 11;

QY 65 LSGVQCFSRYPDHMKQHDF-----FKSAPEGYVQE-----RTI 99
Db 427 IGVGESGISYQASVKQDNFLGTGAAVSTAGTKNDYGTSVNLGYTEPYTKDGVSLGGNV 486
QY 100 FFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNI---LGH-----KLEYNYS 148
Db 487 FFEYDNSKSDTSSNYKRTTYGNSVTL-GFPVNNNSYVVGIGHTYKNISNFALEYN--- 542
QY 149 HNVYIMADKQK-NGIKVNFKIRHNIEDGSGVQLADHYQ-----NTPIGDGPVLL 196
Db 543 RNLYIQSMKFKNGIKTN-----DFDFSGWYNLSNLRGYFPTKGVKASLG-GRVTI 593
QY 197 P--DNHLYLSTOSALS KDPNEKRDMVLLGCFVTAAGITLG 233
Db 594 PGSDNKYKLSADVQGFPLDRDLHLVWVSASAGYANG 632

RESULT 11
F64102
protective surface antigen D-15 - Haemophilus influenzae (strain Rd KW20)
C/Species: Haemophilus influenzae
C/Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 18-Sep-1998
C/Accession: F64102
R/Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.;
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.
A/Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A/Reference number: A64000; MUID:95350630; PMID:7542800
A/Accession: F64102
```



[illegible]

Search completed: June 21, 2004, 16:01:56  
Job time : 11.3333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:42:24 ; Search time 6.44444 Seconds  
(without alignments)  
1931.085 Million cell updates/sec

Title: US-09-887-784-64V

Perfect score: 1273

Sequence: 1 MVSKEELFTGVVPIVLVD.....VLLGFTVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1242	97.6	238	1	GFP_AEQVI
2	93.5	7.3	861	1	SYL_HAEIN
3	91.5	7.2	879	1	SYL_XYLFA
4	89.5	7.0	860	1	SYL_XYLFA
5	89.5	7.0	860	1	SYL_SALTI
6	88.5	7.0	879	1	SYL_SALTY
7	86.5	6.8	533	1	CP51_CANGA
8	86.5	6.8	788	1	DPOL_HPEHE
9	86.5	6.8	795	1	D152_HAEIN
10	86.5	6.8	797	1	D151_HAEIN
11	86.5	6.8	859	1	SYL_SHEON
12	86	6.8	357	1	TRMA_CAWJE
13	86	6.8	461	1	PSBC_CVAPA
14	85.5	6.7	538	1	GRBE_RAT
15	85.5	6.7	793	1	IT33_MESAU
16	85.5	6.7	886	1	IT33_MESAU
17	85.5	6.7	941	1	GUN_EACS6
18	85.5	6.7	1603	1	VID4_CABEL
19	85	6.7	336	1	YD48_METJA
20	85	6.7	439	1	AMPA_WIGBR
21	84.5	6.6	501	1	PEPF_WICPU
22	84.5	6.6	613	1	PEPF_WICPU
23	84	6.6	504	1	YC03_KLEPN
24	84	6.6	682	1	PRC_ECOLI
25	83.5	6.6	1164	1	BAG_STRAG
26	82.5	6.5	860	1	SYL_ECO57
27	82.5	6.5	860	1	SYL_ECOL6
28	82.5	6.5	860	1	SYL_ECOLI
29	82.5	6.5	2222	1	DPOE_YEAST
30	82	6.4	353	1	HIS7_BUCAI
31	82	6.4	689	1	AC2L_HUMAN
32	82	6.4	752	1	NEC1_RAT
33	82	6.4	774	1	AMY2_SCHPO

RESULT 1				
GFP_AEQVI	STANDARD;	PRT;	238	AA.
ID	GFP_AEQVI			
AC	P42212; Q17104; Q27903;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Green fluorescent protein.			
GN	GFP.			
OS	Aequorea victoria (Jellyfish).			
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;			
OC	Aequoreidae; Aequorea.			
OX	NCBI_TaxID=6100;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=92175527; PubMed=1347277;			
RA	Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,			
RA	Cormier M.J.;			
RT	"Primary structure of the Aequorea victoria green-fluorescent			
RT	protein."			
RL	Gene 111:229-233(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94185810; PubMed=8137953;			
RA	Inouye S., Tsuji F.I.;			
RT	"Aequorea green fluorescent protein. Expression of the gene and			
RT	fluorescence characteristics of the recombinant protein."			
RL	FEBS Lett. 341:277-280(1994).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97299832; PubMed=9154981;			
RA	Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;			
RT	"Enhanced expression in tobacco of the gene encoding green fluorescent			
RT	protein by modification of its codon usage."			
RL	Plant Mol. Biol. 33:989-999(1997).			
RN	[4]			
RP	CHROMOPHORE.			
RX	MEDLINE=93192221; PubMed=8448132;			
RA	Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;			
RT	"Chemical structure of the hexapeptide chromophore of the Aequorea			
RT	green-fluorescent protein."			
RL	Biochemistry 32:1212-1218(1993).			
RN	[5]			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=96355665; PubMed=8703075;			
RA	Ormo M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,			
RT	"Crystal structure of the Aequorea victoria green fluorescent			
RT	protein."			
RL	Science 273:1392-1395(1996).			
RN	[6]			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=98294543; PubMed=9631087;			
RA	Yang F., Moss L.G., Phillips G.N. Jr.;			
RT	"The molecular structure of green fluorescent protein."			
RL	Nat. Biotechnol. 14:1246-1251(1996).			

P49052 bacillus li  
Q7vni0 haemophilus  
P53621 homo sapien  
Q25443 helicobacte  
Q9ugn5 homo sapien  
Q00116 homo sapien  
Q8zdf8 yersinia pe  
P91679 drosophila  
P03162 duck hepati  
Q9pkt7 chlamydia m  
P94126 azorhizobiu  
Q63416 rattus norv

ALIGNMENTS

[7] X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.  
MEDLINE=98455509; PubMed=9782051;  
Wachter R.M., Eislinger M.A., Kallio K., Hanson G.T., Remington S.J.;  
"Structural basis of spectral shifts in the yellow-emission variants  
of green fluorescent protein.";  
Structure 6:1267-1277(1998).  
[8]  
X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
MEDLINE=99238303; PubMed=10220315;  
Eislinger M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J.;  
"Structural and spectral response of green fluorescent protein  
variants to changes in pH";  
Biochemistry 38:5296-5301(1999).  
-1- FUNCTION: Energy-transfer acceptor. Its role is to transduce the  
blue chemiluminescence of the protein aequorin into green  
fluorescent light by energy transfer. Fluoresces in vivo upon  
receiving energy from the Ca(2+)-activated photoprotein aequorin.  
Absorbs light maximally at 395 nm and exhibits a smaller  
absorbance peak at 470 nm. The fluorescence emission spectrum  
peaks at 509 nm with a shoulder at 540 nm.  
-1- SUBUNIT: Monomer.  
-1- TISSUE SPECIFICITY: Photocytes.  
-1- PTM: Contains a covalently attached chromophore, which is composed  
of modified amino acid residues. The chromophore is formed upon  
cyclization of the residues Ser-dehydroTy-Gly.  
-1- BIOTECHNOLOGY: Has become a useful and ubiquitous tool for making  
chimeric proteins of GFP linked to other proteins where it  
functions as a fluorescent protein tag. GFP tolerates N- and C-  
terminal fusion to a broad variety of proteins. It has been  
expressed in bacteria, yeast, slime mold, plants, Drosophila,  
zebrafish, and in mammalian cells. As a noninvasive fluorescent  
marker in living cells, it allows for a wide range of applications  
where it may function as a cell lineage tracer, reporter of gene  
expression, or as a measure of protein-protein interactions.  
-1- DATABASE: NAMB-Protein Spotlight;  
NOTE=Issue 11 of June 2001;  
WWW="http://www.expasy.org/spotlight/articles/sptlt011.html".  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M62654; AAA27722.1; -;  
EMBL; M62653; AAA27721.1; -;  
EMBL; J29345; AAA58246.1; -;  
EMBL; X96418; CAA65278.1; -;  
FIR; J30692; JQ1514.  
DR PDB; 1B9C; 17-NOV-00.  
DR PDB; 1BFP; 07-JUL-97.  
DR PDB; 1CAF; 14-JUN-00.  
DR PDB; 1EMA; 08-NOV-96.  
DR PDB; 1EMB; 16-JUN-97.  
DR PDB; 1EMC; 20-AUG-97.  
DR PDB; 1EME; 20-AUG-97.  
DR PDB; 1EMF; 20-AUG-97.  
DR PDB; 1EMG; 12-MAY-99.  
DR PDB; 1EMK; 20-AUG-97.  
DR PDB; 1EML; 20-AUG-97.  
DR PDB; 1EMW; 20-AUG-97.  
DR PDB; 1F09; 17-NOV-00.  
DR PDB; 1FOB; 17-NOV-00.  
DR PDB; 1GFL; 11-JAN-97.  
DR PDB; 1HCT; 15-JAN-02.  
DR PDB; 1HUY; 04-JUL-01.  
DR PDB; 1JBZ; 07-JAN-03.  
DR PDB; 1JP5; 28-AUG-02.  
DR PDB; 1KVP; 10-APR-02.

DR PDB; 1KYR; 10-APR-02.  
DR PDB; 1KYS; 10-APR-02.  
DR PDB; 1YFP; 28-OCT-98.  
DR PDB; 2EMD; 20-AUG-97.  
DR PDB; 2EMN; 20-AUG-97.  
DR PDB; 2EMO; 20-AUG-97.  
DR InterPro: IPR009017; GFP like.  
DR InterPro: IPR000786; Green\_fl\_protein.  
DR Pfam: PF01353; GFP; 1.  
DR PRINTS; PD01229; GFLUORESCENT.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
KW Luminescence; 3D-structure.  
FT CROSSLINK 65 67 5-imidazolinone (Ser-Gly).  
FT MOD\_RES 66 66 2,3-DIDEHYDROTYROSINE.  
FT VARIANT 100 100 F -> Y.  
FT VARIANT 108 108 T -> S.  
FT VARIANT 141 141 L -> M.  
FT VARIANT 219 219 V -> I.  
FT CONFLICT 2 2 S -> G (IN REF. 3).  
FT CONFLICT 25 25 H -> Q (IN REF. 2).  
FT CONFLICT 80 80 Q -> R (IN REF. 3).  
FT CONFLICT 157 157 Q -> P (IN REF. 2).  
FT CONFLICT 172 172 E -> K (IN REF. 2).  
FT HELIX 4 8  
FT STRAND 12 22  
FT TURN 23 24  
FT STRAND 25 36  
FT TURN 37 40  
FT STRAND 41 48  
FT TURN 49 50  
FT TURN 57 60  
FT HELIX 61 63  
FT TURN 69 71  
FT STRAND 73 73  
FT HELIX 76 81  
FT HELIX 83 86  
FT TURN 87 90  
FT STRAND 92 100  
FT TURN 101 102  
FT STRAND 105 115  
FT TURN 116 117  
FT STRAND 118 128  
FT TURN 132 133  
FT TURN 135 139  
FT STRAND 141 141  
FT STRAND 148 155  
FT TURN 156 159  
FT STRAND 160 171  
FT TURN 172 173  
FT STRAND 176 187  
FT STRAND 199 208  
FT TURN 211 212  
FT STRAND 217 227  
SQ SEQUENCE 238 AA; 26886 MW; EA5A6F21FBFB6E05 CRC64;  
Query Match 97.6%; Score 1242; DB 1; Length 238;  
Best Local Similarity 97.9%; Pred. No. 1.7e-95;  
Matches 233; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 2 VSKGELFTGVVPIVLVDGVDVNGHKFSVSGEGGDATYGLTLKFICTTGKLPVWPPTL 61  
Db :|||||  
1 MSKGELFTGVVPIVLVDGVDVNGHKFSVSGEGGDATYGLTLKFICTTGKLPVWPPTL 60  
QY 62 VTVLVSVQCFRSYRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121  
Db |||  
61 VTTFSYVQCFRSYRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120  
QY 122 NRIELKGIIDFKEDGNILGHLKLEYNVNSHNVIYIMADQKNGIKVNFKIRNIEDGSYQLAD 181  
Db |||||  
121 NRIELKGIIDFKEDGNILGHLKLEYNVNSHNVIYIMADQKNGIKVNFKIRNIEDGSYQLAD 180  
QY 182 HYQONTPIGDGPEVLLPDNHYLSTQSALSADPNKRDHMLVLLGFVTAAGITLGMDELYK 239  
|||||

```

Db      181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLBFVTAAGTTHGMDELYK 238
RESULT 2
SYL_HAEIN
ID SYL_HAEIN STANDARD; PRT; 861 AA.
AC P43827;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--trna ligase) (Leurs).
GN LEUS OR HI0921.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + trna(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32774; AAC22581.1; -
CC PIR; H64102; H64102.
CC TIGR; HI0921; -
CC HAMAP; MF 00049; -; 1.
CC InterPro; IPR002302; Leu-trnaSynth1a.
CC InterPro; IPR002300; trna-synt_1a.
CC InterPro; IPR001412; trna-synt_1.
CC InterPro; IPR009008; ValRS_fleRS_edit.
CC Pfam; PF00133; trna-synt 1; 1.
CC PRINTS; P000985; TRNASYNTH1L.
CC TIGRFAMs; TIGR00396; leuS_bact; 1.
CC PROSITE; PS00178; AA TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
FT Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY)
SQ SEQUENCE 861 AA; 97750 MW; EB93304F6B4C8FB7 CRC64;

Query Match 7.3%; Score 93.5; DB 1; Length 861;
Best Local Similarity 24.1%; Pred. No. 4.5;
Matches 46; Conservative 27; Mismatches 59; Gaps 11;

QY 50 TTGKLVPVPTLVTVLSYGVCFSRYPDHMKQHDFFKSAWPEGYQVRIFFKD----- 103
Db      314 TGDKLPV-WVANFVLMHYGTGVAWVPAH-DQRDF-----EFAQKYSIPIKQVIAPLA 364

```

```

QY      104 DQNYKTRAEVKEGDTLVNRIELKGLIDFKEDGNILGHKLEYNYNHNVIMADK-QKNGI 162
Db      365 DEEIDLTKQAFVEHGLKLVNSDFDGKNF--DGAFNG-----IADKLEKLGV 408
QY      163 ---KYNFKIRH-----NIEDGSVOLADHYQQNTPIGDGPVLLPDNHYL- 202
Db      409 GKRQVNYRLRDWGVSRQRYGWGAPIPMLTLIENGDVPA-----FMEDLPILPEDVMD 461
QY      203 STQSALSADPN 213
Db      462 GVKSPINADPN 472
RESULT 3
SYL_XYLFA
ID SYL_XYLFA STANDARD; PRT; 879 AA.
AC Q9P88;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--trna ligase) (Leurs).
GN LEUS OR XF2176.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=9a5G; PubMed=10910347;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Bartos M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Cartaro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorzy H.,
RA Facincini A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hobeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite J.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.P., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Souza A.P., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vialada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + trna(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF004031; AAF84975.1; ALT_INIT.
CC HAMAP; MF_00049; -; 1.

```





```
CC -I- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE008725; AAL19599.1; -.
CC STYGene; SG77777; leus.
CC HAMAP; MF_00049; -. 1.
CC InterPro; IPR002302; Leu-trNASynt1a.
CC InterPro; IPR002300; tRNA-synt_1a.
CC InterPro; IPR001412; tRNA-synt_1.
CC InterPro; IPR009008; ValRS IleRS_edit.
CC Pfam; PF00133; tRNA-synt_1_1.
CC PRINTS; PR00985; TRNASYNTHLEU.
CC TIGRFAMs; TIGR00396; leus bact; 1.
CC PROSITE; PS00178; AA TRNA LIGASE I; 1.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC SITE 42 52 "HIGH" REGION.
CC SITE 619 623 "KMSKS" REGION.
CC BINDING 622 622 ATP (BY SIMILARITY).
CC SEQUENCE 860 AA; 96985 MW; D5003584DFECCAB6 CRC64;
CC
CC Query Match 7.0%; Score 89.5; DB 1; Length 860;
CC Best Local Similarity 23.3%; Pred. No. 9.6;
CC Matches 42; Conservative 22; Mismatches 77; Indels 39; Gaps 7;
CC
CC QY 50 TTGKLPVPPTLTVLSYGVQCFSRYPDHMKQHDFFKSAFMEGVQERTIFFKDDGNK 109
CC Db 314 TGEIIPV-WAANFVLMYGTGAVMVGPH-DQRD-YEFASKYGLTIKPVLAADGSEPD 370
CC
CC QY 110 RAEVKFEGDTLVNRIELKIDFKEDGNILGHKLEYNHNVIMADKQNGIKVNFKR 169
CC Db 371 SEQALTEKGVLFNSGFDGLAFAFNAIDKL-----AEKGGERKVNRYLR 418
CC
CC QY 170 H-----NIEDGSVOLADHYQNTPIGDGPVLLPDNHYL-STQSALSKDP 212
CC Db 419 DWGVSRQRWGAIPMWTLEDGT-----LPTPDQLPVLLPEVDVMDGITSPKADP 471
CC
CC RESULT 6
CC SYL_XYLF
CC ID SYL_XYLF STANDARD; PRT; 879 AA.
CC AC Q87C65;
CC DT 10-OCT-2003 (Rel. 42, Created)
CC DT 10-OCT-2003 (Rel. 42, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE Leucyl-tRNA synthetase (SC 6.1.1.4) (Leucine-tRNA ligase) (LeuRS).
CC GN LEUS OR PD1230.
CC OS Xylella fastidiosa (strain Temecula / ATCC 700964).
CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
CC OC Xanthomonadaceae; Xylella.
CC OX NCBI_TaxID=183190;
CC RN [1]_
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=22421331; PubMed=12533478;
CC RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
CC RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
CC RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
CC RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,
CC RA Carlier H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
CC RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,
CC RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
CC RA Baia G.S., Blanco S.R., Brito M.S., Cannavan P.S., Celestino A.V.,
CC RA da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
```

```
RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Saasaki F.T., Sena J.A.D.,
RA de Souza A.A., Truffi D., Teukumo F., Yanai G.M., Zaros L.G.,
RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA Kitajima J.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of Xylella
RT fastidiosa."
RL J. Bacteriol. 185:1018-1026(2003).
CC -I- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE012557; AAO29080.1; ALT_INIT.
CC HAMAP; MF_00049; -. 1.
CC InterPro; IPR002302; Leu-trNASynt1a.
CC InterPro; IPR002300; tRNA-synt_1a.
CC InterPro; IPR001412; tRNA-synt_1.
CC InterPro; IPR009008; ValRS IleRS_edit.
CC Pfam; PF00133; tRNA-synt_1_1.
CC PRINTS; PR00985; TRNASYNTHLEU.
CC TIGRFAMs; TIGR00396; leus bact; 1.
CC PROSITE; PS00178; AA TRNA LIGASE I; 1.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC SITE 45 55 "HIGH" REGION.
CC SITE 637 641 "KMSKS" REGION.
CC BINDING 640 640 ATP (BY SIMILARITY).
CC SEQUENCE 879 AA; 99823 MW; 4C3EE01B8FDC497E CRC64;
CC
CC Query Match 7.0%; Score 88.5; DB 1; Length 879;
CC Best Local Similarity 22.2%; Pred. No. 12;
CC Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;
CC
CC QY 50 TTGKLPVPPTLTVLSYGVQCFSRYPDHMKQHDFFKSAFMEGVQERTIFFKDDGN 107
CC Db 321 TNEQLPV-WVANFVLMYGTGAVMVGPHDQDQEF--ANKYGLPIRQVIALKEPKN 377
CC
CC QY 108 -----KTRAEVKFEGDTLVNRIELKIDFKEDGNILGHKLEYNHNVYI 153
CC Db 378 STWEPDVMRDWYADKTR---EFE--LINSABFDGLDYQCAFEVLAERFE----- 421
CC
CC QY 154 MADKQKNG-IKYNFKIRHNIEDGSVOLADHYQNTPI-----GDGPVLLPDN 199
CC Db 422 ---RQGRQRRVNYRLR---DWGVSQRQRYGCGPIVYICPTCGAVPVENQLPVLP 474
CC
CC QY 200 -HYLSTQSALSADPNKPKR 216
CC Db 475 VAFSGTGTSPKTDPEWRK 492
CC
CC RESULT 7
CC CP51_CANGA
CC ID CP51_CANGA STANDARD; PRT; 533 AA.
CC AC F50859; Q02312;
CC DT 01-OCT-1996 (Rel. 34, Created)
CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE Cytochrome P450 51 (SC 1.14.13.70) (CYP1) (P450-LIA1) (Sterol 14-
CC DE alpha-demethylase) (lanosterol 14-alpha demethylase) (P450-14DM).
CC GN ERG11 OR CYP51.
CC OS Candida glabrata (Yeast) (Torulopsis glabrata).
CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
CC OX NCBI_TaxID=5478;
```

```

RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=2001-L5;
RX  MEDLINE=96161286; PubMed=8593007;
RA  Geber A., Hitchcock C.A., Swartz J.B., Pullen F.S., Marsden K.E.,
RA  Kwon-Chung K.J., Bennett J.E.;
RT  "Deletion of the Candida glabrata ERG3 and ERG11 genes: effect on cell
RT  viability, cell growth, sterol composition, and antifungal
RT  susceptibility.";
RL  Antimicrob. Agents Chemother. 39:2708-2717(1995).
RN  [2]
RP  SEQUENCE OF 60-473 FROM N.A.
RC  STRAIN=ATCC 2001;
RX  MEDLINE=95081364; PubMed=7989540;
RA  Burgener-Kairuz P., Zuber J.P., Jaunin P., Buchman T.G., Bille J.,
RA  Rossier M.;
RT  "Rapid detection and identification of Candida albicans and
RT  Torulopsis (Candida) glabrata in clinical specimens by
RT  species-specific nested PCR amplification of a cytochrome P-450
RT  lanosterol-alpha-demethylase (LAL1) gene fragment.";
RL  J. Clin. Microbiol. 32:1902-1907(1994).
CC  -1- FUNCTION: Catalyzes C14-demethylation of lanosterol which is
CC  critical for ergosterol biosynthesis. It transforms lanosterol
CC  into 4,4'-dimethyl cholesterol-8,14,24-triene-3-beta-ol (By
CC  similarity).
CC  -1- CATALYTIC ACTIVITY: Obtusifoliol + 3 O(2) + 3 NADPH = 4-alpha-
CC  methyl-5-alpha-ergosta-8,14,24(28)-trien-3-beta-ol + formate + 3
CC  NADP(+) + 3 H(2)O.
CC  -1- PATHWAY: Ergosterol biosynthesis.
CC  -1- SIMILARITY: Belongs to the cytochrome P450 family.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
DR  EMBL; U40389; AAB02329.1; -.
DR  EMBL; S75389; AAB32679.1; -.
DR  InterPro; IPR001128; Cytochrome_P450.
DR  Pfam; PF00067; P450.1.
DR  PRINTS; PR00385; P450.
DR  PROSITE; PS00086; CYTOCHROME_P450; 1.
KW  Electron transport; Oxidoreductase; Monooxygenase; Membrane; Heme;
KW  Sterol biosynthesis; NADP.
FT  METAL 472 472 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT  CONFLICT 64 64 I -> M (IN REF. 2).
FT  CONFLICT 473 473 I -> T (IN REF. 2).
SQ  SEQUENCE 533 AA; 61305 MW; A0506C17507E6EF7 CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 533;
Best Local Similarity 21.8%; Pred. No. 9.6;
Matches 45; Conservative 32; Mismatches 80; Indels 49; Gaps 9;

QY 25 GHKFSVS---GEGEGATYCKILKICITCKLPVWPVLTVLVLSGVQCFSRYPDH--M 79
Db 109 GHEFIFNAKLADVSAAEAYSHLT-----TPV-----FGKGYIDCPNHRLM 149
QY 30 KQHDFFKSAM-PRGYV-----QERTIFFKDDGNKYKTRAEVKFEGDITLVNRIELKGIDF 131
Db 150 EQKKFKGALTKFAFVRYVPLIAEEYKYFRNSKNFKINENSGIVDVVVSQPEM--TIF 207
QY 132 KEDGNILGHKLEYNHSHVNIWADQKNGIKYFKIRNIHEDGVSQVLADHYQNTPIGD 191
Db 208 TASRSLGLKEMRDKLDTDFAYLSLDLKGFTPINF-VFENLPLEHYVKRDKRDAQAITS--- 263
QY 132 GPVLLPDNHYLSTQSALSKDPNEKR 217
Db 254 -----GTYSMLIKERREKND 278
```

```

RESULT 8
DPOL_HPBHE STANDARD; PRT; 788 AA.
AC P13846;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE P protein [Includes: DNA-directed DNA polymerase (EC 2.7.7.7); RNA-
DE directed DNA polymerase (EC 2.7.7.49); Ribonuclease H (EC 3.1.26.4)].
GN P.
OS Heron hepatitis b virus.
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=28300;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88333160; PubMed=3418788;
RA Sprengel R., Kaleta E.F., Will H.;
RT "Isolation and characterization of a hepatitis B virus endemic in
RT herons.";
RL J. Virol. 62:3832-3839(1988).
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA}[N].
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
CC phosphomonoester.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR  EMBL; M22056; AAA45738.1; -.
DR  PIR; A30082; JDVLHH.
DR  InterPro; IPR001462; DNAPol_viral_C.
DR  InterPro; IPR000201; DNAPol_viral_N.
DR  InterPro; IPR000477; RVTse.
DR  Pfam; PF00336; DNA_pol_viral_C; 1.
DR  Pfam; PF00242; DNA_pol_viral_N; 1.
DR  Pfam; PF00078; rvt; 1.
DR  ProDom; PD000814; DNAPol_viral_C; 1.
DR  Transfaser; RNA-directed DNA polymerase; DNA-directed DNA polymerase;
KW  Hydrolase; Nuclease; Endonuclease; DNA replication; DNA-binding.
SQ  SEQUENCE 788 AA; 90070 MW; FB44F38F75EADF44 CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 788;
Best Local Similarity 19.7%; Pred. No. 15;
Matches 40; Conservative 32; Mismatches 66; Indels 65; Gaps 10;

QY 58 WPTLVTVLSYGVQCFSRYPDHMKQH-----DFKSAPEGYVQERT----IPFKDDGNKY 108
Db 139 WPKSISYLPVHSGVKPKYPEFOONHESLVNDYLNKLFEAGILYKRVSKHLVTFK--GPVF 196
QY 109 T-----RAEVKFECDITLVNRIELKGIDFEDGDNILGHKLEYNHSHVYIM 154
Db 197 TWEQKHLVPOQHGAVSSKINDQESRRRIITATSRKNDSSRI-----FGAHN---- 245
QY 155 ADQKNGIKYFKIRNIHEDGVSQVLADHYQ-----QNTPIGDGPVLL--PNHYL 202
Db 246 -----NGRKISY---HSTRDGSRLSGRTSDPTRSGALAGGSDTPIGSGTAHPSTHHV 297
QY 203 STQ-----SALSKDPNEKR 216
Db 298 DRRRRQKGGVQLQAISREPSETR 320

RESULT 9
D152_HAEIN
ID D152_HAEIN STANDARD; PRT; 795 AA.
AC P44935;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
```

RESULT 10
D151_HAEI
ID_D151
AC P460
DT 01-N
DT 01-N
DT 16-O
DE PROT

```
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (Leurs).
GN LEUS OR S01174.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1.
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Teapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Unayam L.A., White O., Wolf A.M.,
RA Vamathavan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uitterlinden T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123 (2002).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB015561; AN54244.1; -.
DR TIGR; S01174; -.
DR HAMAP; MF 00049; -; 1.
DR InterPro; IPR002302; Leu-tRNA synthetase.
DR InterPro; IPR002300; tRNA-synt 1a.
DR InterPro; IPR001412; tRNA-synt 1.
DR InterPro; IPR009008; ValRS-like edit.
DR Pfam; PF00133; tRNA-synt 1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRFAMs; TIGR00396; leuS bact; 1.
DR PROSITE; PS00178; AA TRNA LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 618 622 "KMSKS" REGION.
FT BINDING 621 621 ATP (BY SIMILARITY).
FT SEQUENCE 859 AA; 96827 MW; C78D6209DFB6CA17 CRC64;
Query Match 6.8%; Score 86.5; DB 1; Length 859;
Best Local Similarity 23.6%; Pred. No. 17;
Matches 50; Conservative 30; Mismatches 95; Indels 37; Gaps 9;
QY 18 ELDDGVNGHKFSVSGEGE-----GDATVGLKTLKFICTGKLPVPTLVTLVSGVQ 70
Db 277 ELAAFDCEKNSTSAELATMEKRGVAT-GLYAIHPI--TGQVPIWAFVIMNYGTG 333
QY 71 CFSRYPDH-MKHDFPKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLVNRIELKGI 129
Db 334 AVMSVPGHQDQDYEFAK----KVHLPTEAVIKPAEGDLIDISEAAYTEKGLIFNSGEFDGL 389
QY 130 DKEDGNILGHKLEYNNSHNHYIMADKQNGIKVNFKIRH-----NIEDGSGVQLA 180
Db 390 DFDGAFNIVANKL-----VABGKGR--QVNYRLRDMGVSRQRYWGAPIPMVTLA 437
QY 131 DHYQONTPTGDGVPVLPDNLHNL-STQSALSKD 211
Db 438 DGTVIPTPDQLPVLIPEDVMDIGSPIKAD 469
```

```
RESULT 12
TRNA_CAMJE STANDARD; PRT; 357 AA.
AC Q9P92; 2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE tRNA (Uracil-5-)-methyltransferase (EC 2.1.1.35) (tRNA(M-5-U54)-
DE methyltransferase) (RUMT).
GN TRMA OR CJ0831C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteriaceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
CC -1- FUNCTION: Catalyzes the formation of 5-methyl-uridine at position
CC 54 (M-5-U54) in all tRNA (By similarity).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
CC homocysteine + tRNA containing thymine.
CC -1- SIMILARITY: Belongs to the RNA MSU methyltransferase family. Trna
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL139076; CAB73096.1; -.
DR HAMAP; MF 01011; -; 1.
DR InterPro; IPR000051; SAM bind.
DR InterPro; IPR001566; TrmA.
DR PROSITE; PS01230; TRMA_1; 1.
DR PROSITE; PS01231; TRMA_2; FALSE NEG.
DR TRANSFERASE; Methyltransferase; tRNA processing; Complete proteome.
FT DOMAIN 207 213 S-ADENOSYLMETHIONINE BINDING (BY
FT SIMILARITY).
FT ACT_SITE 315 315 BY SIMILARITY.
FT SEQUENCE 357 AA; 42276 MW; CEC5328347CEB497 CRC64;
Query Match 6.8%; Score 86; DB 1; Length 357;
Best Local Similarity 24.8%; Pred. No. 6.5;
Matches 30; Conservative 18; Mismatches 39; Indels 34; Gaps 5;
QY 80 KQHDFPKSAMPEGYVQERTIFFKDDGNKYKTRAEVKF--EGDITLV-----NRIELKG 128
Db 14 EKHSITKIKYKFFYTKDFKFLFASKDKHKYTRAEISFYHENDTLFYAMFPDKSKKYLIEY 73
QY 129 IDPKED-----GNILGHKLEYNNSHNHYIMADKQNGIKVNFKIRHIE 173
Db 74 LDFADEKICAFPRLLLEYLRQDNKLEKL-----FGVEFLTQKE--LSITLLYHKNIE 125
QY 174 D 174
Db 126 D 126
RESULT 13
PSBC_CYAPA
```

```

ID PSBC_CYPAPA STANDARD; PRT; 461 AA.
AC P48104;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Photosystem II 44 kDa reaction center protein (p6 protein) (CP43).
GN PSBC.
OS Cyanophora paradoxa.
OG Cyanelle.
OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OX NCBI_TaxID=2762;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
RA Bryant D.A.;
RT "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa.";
RL Plant Mol. Biol. Rep. 13:327-332(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
RA Farley J.Y., Schlachter W.M., Chung S., Newmann-Spallart C.,
RA Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
RT the genetic complexity of a primitive plastid.";
RL (in) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,
RL Schwemmler W. (eds.);
RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
RL (1997).
CC -!- FUNCTION: The 43 kDa protein (p6) is a component of the core of
CC photosystem II. It is a chlorophyll binding protein.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Cyanelle
CC thylakoid membrane.
CC -!- SIMILARITY: Belongs to the psbB / psbC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U30821; AAA61279.1; -.
CC PIR; T06936; T06936. Photo44.
CC InterPro; IPR005869; Photo44.
CC InterPro; IPR000932; PSIIProt.
CC Pfam; PF00421; PSI1; 1.
CC TIGRPFAMs; TIGR01153; psbC; 1.
CC Photosynthesis; Photosystem II; Thylakoid; Chlorophyll; Cyanelle;
CC Transmembrane.
CC SEQUENCE 461 AA; 50402 MW; 62664E239E13B9C1 CRC64;

Query Match 6.8%; Score 86; DB 1; Length 461;
Best Local Similarity 21.8%; Pred. No. 8.9;
Matches 55; Conservative 31; Mismatches 88; Indels 78; Gaps 12;

QY 25 GHKFSVSGEGDGYTKLTKPCTTGKL-----PVPWPLTVV-----LSYGV---69
DB 208 GDCWIVSDNMEDIIGGHIWLAFCIIGGVWHLTKPFSWARRALVMSGEAVLSYSLAAL 267
QY 70 -----QCFSRYPDHMKQHDPEKFSAMPE-GYVQERTIFFKOD-----GN 106
DB 268 ALMGFTIANCFVFNNTAYPSEFPPTGPEASQAQFTFLVDRQRLGANVGSAGQPTGLGK 327
QY 107 YKTRA---EVKFEGLTLVNRIELKIDGDFKDNGLKHLEYNYNHNHYIMADKQKNGIK 163
DB 328 YLWRSFSGEIIIFGETM-----RFDWTRAPWLEPLRGANGLD 364
QY 164 VNFKIRHNIEDSGVLADHYQQNTPIGD---GPVLLPDN---HYLSTQALSCKDPNEKD 217
DB 365 LT-KIKYDIQPWQERRAAEYMTHTAPLGSLNSVSGVATEINSVNVSPRSWLS-----TS 417
```

```

QY 218 HMVLGFTVTAAG 229
DB 418 HFV-LGFFLFIG 428

RESULT 14
GRBE_RAT
ID -GRBE_RAT STANDARD; PRT; 538 AA.
AC O88900;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Growth factor receptor-bound protein 14 (GRB14 adapter protein).
GN GRB14.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar.
RX MEDLINE=98421528; PubMed=9748281;
RA Kasus-Jacobi A., Perdereau D., Auzan C., Clauser E., van Obberghen E.,
RA Mauvais-Jarvis F., Girard J., Burnol A.-F.;
RT "Identification of the rat adapter Grb14 as an inhibitor of insulin
RT actions.";
RL J. Biol. Chem. 273:26026-26035(1998).
CC -!- FUNCTION: Interacts with the cytoplasmic domain of the
CC autophosphorylated insulin receptor which is then inhibited. The
CC interaction is mediated by the SH2 domain.
CC -!- SUBUNIT: Binds to the ankyrin repeat region of TNKL via its N-
CC terminus (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and
CC endosomes (By similarity).
CC -!- PTM: Phosphorylated on serine residues (By similarity).
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 1 Ras-associating domain.
CC -!- SIMILARITY: Contains 1 SH2 domain.
CC -!- SIMILARITY: Belongs to the GRB7/10/14 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF076619; AAC61478.1; -.
CC HSSP; P35235; IAYA.
CC InterPro; IPR001849; PH.
CC InterPro; IPR000159; RA domain.
CC InterPro; IPR000980; SH2.
CC Pfam; PF00169; PH; 1.
CC Pfam; PF00788; RA; 1.
CC Pfam; PF00017; SH2; 1.
CC PRINTS; PR00401; SH2DOMAIN.
CC ProDom; PD000093; SH2; 1.
CC SMART; SM00233; PH; 1.
CC SMART; SM00314; RA; 1.
CC SMART; SM00252; SH2; 1.
CC PROSITE; PS50003; PH DOMAIN; 1.
CC PROSITE; PS50200; RA; 1.
CC PROSITE; PS50001; SH2; 1.
CC SH2 domain; Phosphorylation.
FT DOMAIN 104 190
FT DOMAIN 232 340 PH.
FT DOMAIN 437 533 SH2.
SQ SEQUENCE 538 AA; 60592 MW; C6C9037E7868EEF CRC64;

Query Match 6.7%; Score 85.5; DB 1; Length 538;
Best Local Similarity 19.0%; Pred. No. 12;
```

```
Matches 40; Conservative 34; Mismatches 70; Indels 67; Gaps 9;
QY 41 GKLTLLKFICT-TKKLVPWPMTLVTLVLSVGQCFSRY--PDHMKQHDFFKSAMPEGYVQER 97
Db 313 GPRDLKMLCAEDDSRMCHWTAIRLLKYGMQLYQNYMHPQSARSACSSQSVSPMSVSEN 372
QY 98 TIFFKDDGNKYKTRAEVKFEGDTLVNRIELKIDFEDG-----NIL 138
Db 373 SLVAMDFSGQKTRV-IDNPTEALSVAVE-EGLAWRKGCGLRGNHGSPTAPSSAVNMA 430
QY 139 GHKLEYNYNH-----NVYIMADKQKN-----GIKVNF----- 166
Db 431 LHRSPWFHHRISRDEAQQLIITRQGVGVFLVRDSQSNPRTFVLSMHSGQKIKHFIIP 490
QY 167 -----KIRHNIEDGS-----VOLADHYQON 186
Db 491 VEDDGEVFTLDDGHTKFTDLIQLVFEYQLN 521
```

```
RESULT 15
D153 HAEIN
ID D153 HAEIN STANDARD; PRT; 793 AA.
AC Q32629;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=PAK 12085;
RX MEDLINE=97427952; PubMed=9284140;
RA Loomis S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease."
RL Infect. Immun. 65:3701-3707(1997).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: Belongs to the surface antigen D15 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U60834; AAB61977.1;
DR InterPro: IPR000184; Bac_surfAg_D15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 793 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 793 AA; 87511 MW; 51BFDB2036801A14 CRC64;
```

```
Query Match 6.7%; Score 85.5; DB 1; Length 793;
Best Local Similarity 22.4%; Pred. No. 19;
Matches 49; Conservative 28; Mismatches 79; Indels 63; Gaps 11;
```

```
QY 65 LSYGVQCFSRYPDHMKQHDF-----FKSAMPEGYVQE-----RTI 99
Db 427 IGYGTSGISYQTSIKQDNFLGTGAVSIACTKNDYGTSVNLGYTEPYFTKDGVSLGNI 486
QY 100 FFKDDGNKYKTRAEVKFEGDTLVNRIELKIDFEDGNI---LGH-----KLEYNYN 148
Db 187 FFENVDSKSDTSSNYKRTTYGSNVTL-GFPVNNNSYYGLGHTYKNKISNFALEYN--- 542
```

```
QY 149 HNVYIMADKQK-NGIKVNFKIRHNIEDGSVOLADHYQO-----NTPIGDPVLL 196
Db 543 RNLYIQSMKFKONGIKTN-----DFDFSFGWYNYSLNRGYPTKGVKASLG-GRVTI 593
QY 197 P--DNHYLSTQSALSKDPNERKRDHMLLGFTVTAAGITLG 233
Db 594 PGSDNKYYKLSADVQGFYPLDRDRHRVVSAKASAGYANG 632
```

Search completed: June 21, 2004, 15:55:20

Job time : 7.55556 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:45:30 ; Search time 30.6667 Seconds  
(without alignments)  
2458.984 Million cell updates/sec

Title: US-09-887-784-64v  
Perfect score: 1273  
Sequence: 1 MNSGSELFTGVVPIVLVD.....VLIGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL 25:\*
- 1: sp archaea:\*
  - 2: sp bacteria:\*
  - 3: sp fungi:\*
  - 4: sp human:\*
  - 5: sp invertebrate:\*
  - 6: sp mammal:\*
  - 7: sp mhc:\*
  - 8: sp organelle:\*
  - 9: sp phase:\*
  - 10: sp plant:\*
  - 11: sp rodent:\*
  - 12: sp virus:\*
  - 13: sp vertebrate:\*
  - 14: sp unclassified:\*
  - 15: sp virus:\*
  - 16: sp bacteriophage:\*
  - 17: sp archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1238	97.3	238	2 Q8GHE2	Q8ghe2 azotobacter
2	1235	97.0	238	5 Q93125	Q93125 aequorea vi
3	1233	96.9	238	2 Q8GHE4	Q8ghe4 azomonas ag
4	1232	96.8	238	2 Q8GHE3	Q8ghe3 azotobacter
5	1200	94.3	238	5 Q17105	Q17105 aequorea vi
6	1185	93.1	238	5 Q17106	Q17106 aequorea vi
7	1080	84.8	238	5 Q8WTC6	Q8wtc6 aequorea ma
8	1076	84.5	238	5 Q8WTP95	Q8wtp95 aequorea ma
9	1072	84.2	238	5 Q8WTC4	Q8wtc4 aequorea ma
10	1070	84.1	238	5 Q8WTD0	Q8wtcd0 aequorea ma
11	1069	84.0	238	5 Q8WTC8	Q8wtc8 aequorea ma
12	1069	84.0	238	5 Q8WTC9	Q8wtc9 aequorea ma
13	1067	83.8	238	5 Q8WTC7	Q8wtc7 aequorea ma
14	1065	83.7	238	5 Q8WTC5	Q8wtc5 aequorea ma
15	252.5	19.8	225	5 Q95UA7	Q95ua7 montastraea
16	252.5	19.8	225	5 Q720W5	Q720w5 montastraea

17	247	19.4	225	5	Q963F5	Q963f5 montastraea
18	246.5	19.4	225	5	Q720W9	Q720w9 montastraea
19	244.5	19.2	236	5	Q8T6U0	Q8t6u0 dendronepht
20	240	18.9	225	5	Q8I6J8	Q8i6j8 trachyphyll
21	238.5	18.7	286	5	Q9U6Y3	Q9u6y3 clavularia
22	236	18.5	225	5	Q8T5F1	Q8t5f1 montastraea
23	233	18.3	225	5	Q720W4	Q720w4 montastraea
24	232	18.2	224	5	Q8MU48	Q8mu48 montastraea
25	218	17.1	227	5	Q720W6	Q720w6 montastraea
26	218	17.1	234	5	Q720W7	Q720w7 montastraea
27	213	16.7	234	5	Q8T5F2	Q8t5f2 montastraea
28	213	16.7	234	5	Q8MU47	Q8mu47 montastraea
29	212.5	16.7	229	5	Q9U6Y6	Q9u6y6 anemonia ma
30	212.5	16.7	259	5	Q8MMA2	Q8mma2 agaricia fr
31	212	16.7	239	5	Q8MMA1	Q8mma1 agaricia ag
32	210	16.5	227	5	Q962P9	Q962p9 montastraea
33	210	16.5	227	5	Q720W8	Q720w8 montastraea
34	206	16.2	227	5	Q95VT0	Q95vt0 montastraea
35	203.5	16.0	238	5	Q9BLY9	Q9bly9 renilla mue
36	203	15.9	221	5	Q95P04	Q95p04 goniopora t
37	202.5	15.9	232	5	Q9GP15	Q9gpi5 anemonia su
38	200.5	15.8	225	5	Q9U6Y8	Q9u6y8 discosoma s
39	199.5	15.7	214	5	Q86LV7	Q86lv7 meandrina m
40	198.5	15.6	214	5	Q86LV8	Q86lv8 meandrina m
41	198.5	15.6	222	5	Q72168	Q72168 cerianthus
42	198.5	15.6	225	5	Q8T6T9	Q8t6t9 radianthus
43	198.5	15.6	232	5	Q9GZ28	Q9gz28 anemonia su
44	198	15.6	235	5	Q8T5F0	Q8t5f0 scolymia cu
45	196.5	15.4	232	5	Q9U6Y7	Q9u6y7 discosoma s

ALIGNMENTS

RESULT 1

ID	Q8GHE2	PRELIMINARY;	PRT;	238 AA.
AC	Q8GHE2;			
DT	01-MAR-2003 (TrEMBLrel. 23, Created)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)			
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE	Green fluorescence protein.			
GN	2289GFP.			
OS	Azotobacter vinelandii.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;			
OC	Pseudomonadaceae; Azotobacter.			
OX	NCBI_TaxID=354;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=DSM2289;			
RA	Koranyi P., Berenyi M., Burg K.;			
RT	"Occurrence of green fluorescence protein in diazotrophic bacteria			
RT	Azomonas and Azotobacter.";			
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF324408; AAN86140.1; -;			
DR	GO: GO:0006091; P:energy pathways; IEA.			
DR	InterPro; IPR009017; GFP_like.			
DR	InterPro; IPR000786; Green_fl_protein.			
DR	Pfam; PF01353; GFP; 1.			
DR	PRINTS; PR01229; GFP_LUORESCENT.			
DR	ProDom; PD013756; Green fl protein; 1.			
SQ	SEQUENCE 238 AA; 26914 MW; F84840F1F9064018 CRC64;			

Query Match	97.3%;	Score 1238;	DB 2;	Length 238;
Best Local Similarity	97.5%;	Pred. No. 9.3e-97;		
Matches 232;	Conservative	2;	Mismatches 4;	Indels 0; Gaps 0;
Qy	2	VSKGEELFTGVVPIVLVDGNGHKFSVSGEGEDATYGKLTCLKFTCTTGKLPVWPPTL	61	
Db	1	MSGGEELFTGVVPIVLVDGNGHKFSVSGEGEDATYGKLTCLKFTCTTGKLPVWPPTL	60	
Qy	62	VTVLSYGVQCFSRYPDHMKOHDFFKSAPMEGYVQERTIFFKDDGNKYKTRAEVFEFGDTLV	121	

```
Db 61 VTTFSYGVQCFSRYPDHMKRHHDFKSAHPGEGYVQERTIFFKDDGNKYKTRAEVKEFGDTLV 120
Qy 122 NRIELKGIQDKEDGNILGHKLEYNYNHNNYIMADKQKNGIKVNFIRHNIEDGSVOLAD 181
Db 121 NRIELKGIQDKEDGNILGHKLEYNYNHNNYIMADKQKNGIKVNFIRHNIEDGSVOLAD 180
Qy 182 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNKRHDHMLLGFVTAAGITLGMDELK 239
Db 181 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNKRHDHMLLGFVTAAGITLGMDELK 238

RESULT 2
Q93125 PRELIMINARY; PRT; 238 AA.
AC Q93125;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein mutant 3.
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96305137; PubMed=8707053;
RA Cormack B.P., Valdivia R.H., Falkow S.;
RT "FACS-Optimized mutants of the green fluorescent protein (GFP).";
RL Microbiology 0:0-0(1996).
DR EMBL; U73901; AAB18957.1; -.
DR HSP; P42212; IAFP.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP_UOESCENT.
DR PRODOM; PD013756; Green fl protein; 1.
DR SEQUENCE 238 AA; 26840 MW; A28622809A9DEA60 CRC64;

Qy 2 VSKGEELFTGVPIVLVDGNGHKFSVSGEGDGYKLTILKFKICTTGKLPVWPPTL 61
Db 1 MSKGEELFTGVPIVLVDGNGHKFSVSGEGDGYKLTILKFKICTTGKLPVWPPTL 60

Qy 62 VTVLSYGVQCFSRYPDHMKRHHDFKSAHPGEGYVQERTIFFKDDGNKYKTRAEVKEFGDTLV 121
Db 61 VTTFSYGVQCFSRYPDHMKRHHDFKSAHPGEGYVQERTIFFKDDGNKYKTRAEVKEFGDTLV 120

Qy 122 NRIELKGIQDKEDGNILGHKLEYNYNHNNYIMADKQKNGIKVNFIRHNIEDGSVOLAD 181
Db 121 NRIELKGIQDKEDGNILGHKLEYNYNHNNYIMADKQKNGIKVNFIRHNIEDGSVOLAD 180

Qy 182 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNKRHDHMLLGFVTAAGITLGMDELK 239
Db 181 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNKRHDHMLLGFVTAAGITLGMDELK 238

RESULT 3
Q8GHE4 PRELIMINARY; PRT; 238 AA.
AC Q8GHE4
Q8GHE4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Green fluorescent protein.
GN 8SGFP.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
Azomonas and Azotobacter."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324406; AAN86138.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP_UOESCENT.
DR PRODOM; PD013756; Green fl protein; 1.
DR SEQUENCE 238 AA; 26887 MW; E0E1616BD2AF6188 CRC64;

Query Match 96.8%; Score 1232; DB 2; Length 238;
Best Local Similarity 97.1%; Pred. No. 1.7e-96;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
```

```
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescence protein.
GN 375GFP.
OS Azomonas agilis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azomonas.
OX NCBI_TaxID=116849;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
Azomonas and Azotobacter."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324405; AAN86137.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP_UOESCENT.
DR PRODOM; PD013756; Green fl protein; 1.
DR SEQUENCE 238 AA; 26902 MW; 15F9B9C5B4F6B89 CRC64;

Query Match 96.9%; Score 1233; DB 2; Length 238;
Best Local Similarity 97.1%; Pred. No. 2.5e-96;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
```

Qy 2 VSKGEELFTGVPIVLVDGNGHKFSVSGEGDGYKLTILKFKICTTGKLPVWPPTL 61  
Db 1 MSKGEELFTGVPIVLVDGNGHKFSVSGEGDGYKLTILKFKICTTGKLPVWPPTL 60

Qy 62 VTVLSYGVQCFSRYPDHMKRHHDFKSAHPGEGYVQERTIFFKDDGNKYKTRAEVKEFGDTLV 121  
Db 61 VTTFSYGVQCFSRYPDHMKRHHDFKSAHPGEGYVQERTIFFKDDGNKYKTRAEVKEFGDTLV 120

Qy 122 NRIELKGIQDKEDGNILGHKLEYNYNHNNYIMADKQKNGIKVNFIRHNIEDGSVOLAD 181  
Db 121 NRIELKGIQDKEDGNILGHKLEYNYNHNNYIMADKQKNGIKVNFIRHNIEDGSVOLAD 180

Qy 182 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNKRHDHMLLGFVTAAGITLGMDELK 239  
Db 181 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNKRHDHMLLGFVTAAGITLGMDELK 238

RESULT 4  
Q8GHE3 PRELIMINARY; PRT; 238 AA.  
AC Q8GHE3;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Green fluorescence protein.  
GN 8SGFP.  
OS Azotobacter vinelandii.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Azotobacter.  
OX NCBI\_TaxID=354;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Koranyi P., Berenyi M., Burg K.;  
RT "Occurrence of green fluorescence protein in diazotrophic bacteria  
Azomonas and Azotobacter."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324406; AAN86138.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP\_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP\_UOESCENT.
DR PRODOM; PD013756; Green fl protein; 1.
DR SEQUENCE 238 AA; 26887 MW; E0E1616BD2AF6188 CRC64;

Query Match 96.8%; Score 1232; DB 2; Length 238;



```
Best Local Similarity 97.1%; Pred. No. 3e-96;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFTCTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFTCTTGKLPVWPPTL 60

QY 62 VTVLGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKYKFKIRHNIEDGSVQLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKYKFKIRHNIEDGSVQLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNPKRDHMLLVFTAAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNPKRDHMLLVFTAAAGITLGMDELYK 238

RESULT 5
Q17105
ID Q17105 PRELIMINARY; PRT; 238 AA.
AC Q17105;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83959; CAA58789.1; -.
DR PIR; JS0692; J01514.
DR HSSP; P42212; IGFL.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748544 CRC64;

Query Match 94.3%; Score 1200; DB 5; Length 238;
Best Local Similarity 93.7%; Pred. No. 1.5e-93;
Matches 223; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFTCTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFTCTTGKLPVWPPTL 60

QY 62 VTVLGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKYKFKIRHNIEDGSVQLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKYKFKIRHNIEDGSVQLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNPKRDHMLLVFTAAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNPKRDHMLLVFTAAAGITLGMDELYK 238

RESULT 6
Q17106
ID Q17106 PRELIMINARY; PRT; 238 AA.
```

```
Q17106;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83960; CAA58790.1; -.
DR PIR; JS0692; J01514.
DR HSSP; P42212; IBFP.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26867 MW; BD4648262D8EABD4 CRC64;

Query Match 93.1%; Score 1185; DB 5; Length 238;
Best Local Similarity 92.9%; Pred. No. 2.8e-92;
Matches 221; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFTCTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFTCTTGKLPVWPPTL 60

QY 62 VTVLGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKYKFKIRHNIEDGSVQLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKYKFKIRHNIEDGSVQLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNPKRDHMLLVFTAAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNPKRDHMLLVFTAAAGITLGMDELYK 238

RESULT 7
Q8WTC6
ID Q8WTC6 PRELIMINARY; PRT; 238 AA.
AC Q8WTC6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPxm19uv;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RL "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla."
DR EMBL; AF435431; AAL33916.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR Pfam; PF01353; GFP; 1.
```

```
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27015 MW; 688FD75E88926903 CRC64;

Query Match      84.8%; Score 1080; DB 5; Length 238;
Best Local Similarity 81.9%; Pred. No. 2.2e-83;
Matches 195; Conservative 20; Mismatches 23; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVELDGVNKHGKFSVSGEGDGYKLTKEICTTGKLPVWPPTL 61
DB 1 MSGGEELFTGIVPVLIELDGVNKHGKFSVRGEGDGYKLTKEICTTGKLPVWPPTL 60

QY 52 VTLSVGVQCFSRYPDHMKQDFFKSAPEGYQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB 51 VTLSVGIQCFARYPEHMKMNDFFKSAPEGYQERTIFFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRLEKGDIDFKEDGNILGHKLEYNHSHVNYIMADKQKGIKYNFKIRHNIEDGSVOLAD 181
DB 121 NRLEKGMDFKEDGNILGHKLEYNHSHVNYIMPDKANGLKYNFKIRHNIEGGVOLAD 180

QY 132 HYQONTPIGDGPVLLPDNHYLSQTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 131 HYQTNVPLGDGPVLLPINHYLSQTQTAISKDRNETRDHVMVLEFFSACGTHGMDELYK 238

RESULT 8
Q8WP95 PRELIMINARY; PRT; 238 AA.
AC Q8WP95;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFPXM.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPxm161;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL33918.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR0009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match      84.2%; Score 1072; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 1e-82;
Matches 196; Conservative 17; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVELDGVNKHGKFSVSGEGDGYKLTKEICTTGKLPVWPPTL 61
DB 1 MSGGEELFTGVVPIVLVELDGVNKHGKFSVRGEGDGYKLTKEICTTGKLPVWPPTL 60

QY 62 VTLSVGVQCFSRYPDHMKQDFFKSAPEGYQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB 61 VTLSVGIQCFARYPEHMKMNDFFKSAPEGYQERTIFFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRLEKGDIDFKEDGNILGHKLEYNHSHVNYIMADKQKGIKYNFKIRHNIEDGSVOLAD 181
DB 121 NRLEKGMDFKEDGNILGHKLEYNHSHVNYIMPDKANGLKYNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 HYQTNVPLGDGPVLLPINHYLSQTQTAISKDRNETRDHVMVLEFFSACGTHGMDELYK 238

RESULT 10
Q8WTD0 PRELIMINARY; PRT; 238 AA.
AC Q8WTD0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPxm161;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL33918.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR0009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match      84.5%; Score 1076; DB 5; Length 238;
Best Local Similarity 81.5%; Pred. No. 4.7e-83;
Matches 194; Conservative 20; Mismatches 24; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVELDGVNKHGKFSVSGEGDGYKLTKEICTTGKLPVWPPTL 61
DB 1 MSGGEELFTGIVPVLIELDGVNKHGKFSVRGEGDGYKLTKEICTTGKLPVWPPTL 60

QY 52 VTLSVGVQCFSRYPDHMKQDFFKSAPEGYQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB 51 VTLSVGIQCFARYPEHMKMNDFFKSAPEGYQERTIFFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRLEKGDIDFKEDGNILGHKLEYNHSHVNYIMADKQKGIKYNFKIRHNIEDGSVOLAD 181
DB 121 NRLEKGMDFKEDGNILGHKLEYNHSHVNYIMPDKANGLKYNFKIRHNIEGGVOLAD 180

QY 132 HYQONTPIGDGPVLLPDNHYLSQTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 131 HYQTNVPLGDGPVLLPINHYLSQTQTAISKDRNETRDHVMVLEFFSACGTHGMDELYK 238
```

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF435427; AAL33912.1; -;  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR009017; GFP like.  
DR InterPro; IPR000786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PD01229; GFLUORESCENT.  
DR ProDom; PD013756; Green fl protein; 1.  
SQ SEQUENCE 238 AA; 26997 MW; 5F80A192173CB84D CRC64;

Query Match 84.0%; Score 1070; DB 5; Length 238;  
Best Local Similarity 81.1%; Pred. No. 1.8e-82;  
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLELGDVNGHKFVSVEGEGDATYGLTKLKFICTTGKLPVWPPTL 61  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 1 MSKGEELFTGIVPVLELGDVNGHKFVSVEGEGDADYGLKEIKTCTTGKLPVWPPTL 60  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
QY 62 VTVLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 61 VTTLGVIQCFARYPEHMKMNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
QY 122 NRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQKGIKYNFKIRHNIEGSGVOLAD 181  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 121 NRIELKGMDFKEDGNILGHKLEYNNSHNHYIMPDKANNGLKNFKIRHNIEGGVQLAD 180  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
QY 182 HYQONTPIGDGPVLLPDNHYLSQTSALSADPNKRDHMLVGFVTAAGITLGMDELYK 239  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 181 HYQTNVPLGDGPVLLIPINHYLSQTSALSADPNKRDHMLVGFVTAAGITLGMDELYK 238  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 11

Q8WTC8 PRELIMINARY; PRT; 238 AA.  
ID Q8WTC8;  
AC Q8WTC8;  
DT 01-MAR-2002 (T-EMBLrel. 20, Created)  
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
DE Green fluorescent protein.  
GN GFP.  
OS Aequorea macrodactyla.  
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;  
OC Aequoreidae; Aequorea.  
OX NCBI\_TaxID=147615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,  
RA Li S.J., Xia N.S.;  
RT "Colorful mutants of green fluorescent protein from Aequorea  
macrodactyla.";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF435429; AAL33914.1; -;  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR009017; GFP like.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PD01229; GFLUORESCENT.  
DR ProDom; PD013756; Green fl protein; 1.  
SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;

Query Match 84.0%; Score 1069; DB 5; Length 238;  
Best Local Similarity 81.1%; Pred. No. 1.8e-82;  
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLELGDVNGHKFVSVEGEGDATYGLTKLKFICTTGKLPVWPPTL 61  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 1 MSKGEELFTGIVPVLELGDVNGHKFVSVEGEGDADYGLKEIKTCTTGKLPVWPPTL 60  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
QY 62 VTVLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 61 VTTLGVIQCFARYPEHMKMNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 122 NRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQKGIKYNFKIRHNIEGSGVOLAD 181  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 121 NRIELKGMDFKEDGNILGHKLEYNNSHNHYIMPDKANNGLKNFKIRHNIEGGVQLAD 180  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
QY 182 HYQONTPIGDGPVLLPDNHYLSQTSALSADPNKRDHMLVGFVTAAGITLGMDELYK 239  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 181 HYQTNVPLGDGPVLLIPINHYLSQTSALSADPNKRDHMLVGFVTAAGITLGMDELYK 238  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 12

Q8WTC9 PRELIMINARY; PRT; 238 AA.  
ID Q8WTC9;  
AC Q8WTC9;  
DT 01-MAR-2002 (T-EMBLrel. 20, Created)  
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
DE Green fluorescent protein.  
GN GFP.  
OS Aequorea macrodactyla.  
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;  
OC Aequoreidae; Aequorea.  
OX NCBI\_TaxID=147615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,  
RA Li S.J., Xia N.S.;  
RT "Colorful mutants of green fluorescent protein from Aequorea  
macrodactyla.";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF435428; AAL33913.1; -;  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR009017; GFP like.  
DR InterPro; IPR000786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PD01229; GFLUORESCENT.  
DR ProDom; PD013756; Green fl protein; 1.  
SQ SEQUENCE 238 AA; 27031 MW; 5F80A19C19DC584D CRC64;

Query Match 84.0%; Score 1069; DB 5; Length 238;  
Best Local Similarity 81.1%; Pred. No. 1.8e-82;  
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLELGDVNGHKFVSVEGEGDATYGLTKLKFICTTGKLPVWPPTL 61  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 1 MSKGEELFTGIVPVLELGDVNGHKFVSVEGEGDADYGLKEIKTCTTGKLPVWPPTL 60  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
QY 62 VTVLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 61 VTTLGVIQCFARYPEHMKMNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 122 NRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQKGIKYNFKIRHNIEGSGVOLAD 181  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 121 NRIELKGMDFKEDGNILGHKLEYNNSHNHYIMPDKANNGLKNFKIRHNIEGGVQLAD 180  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
QY 182 HYQONTPIGDGPVLLPDNHYLSQTSALSADPNKRDHMLVGFVTAAGITLGMDELYK 239  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 181 HYQTNVPLGDGPVLLIPINHYLSQTSALSADPNKRDHMLVGFVTAAGITLGMDELYK 238  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 13

Q8WTC7 PRELIMINARY; PRT; 238 AA.  
ID Q8WTC7;  
AC Q8WTC7;  
DT 01-MAR-2002 (T-EMBLrel. 20, Created)  
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
DE Green fluorescent protein.  
GN GFP.  
OS Aequorea macrodactyla.  
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;  
OC Aequoreidae; Aequorea.  
OX NCBI\_TaxID=147615;



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:41:49 ; Search time 47.1111 Seconds  
(without alignments)  
1433.395 Million cell updates/sec

Title: US-09-887-784-222A

Perfect score: 1274

Sequence: 1 MVSKGEELFTGVVPLVELD.....VLAFGVTAAAGITLGMDELTK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 29Jan04:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1990s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1269	99.6	239	5 AAE17518	Enhanced
2	1269	99.6	363	6 ABR40352	Human ami
3	1269	99.6	893	4 AAG65781	Amino aci
4	1269	99.6	1132	4 AAG65782	Amino aci
5	1261	99.0	239	5 AAE17517	Enhanced
6	1258	98.7	239	3 AAB22882	Enhanced
7	1258	98.7	239	3 AAY54349	Amino aci
8	1258	98.7	239	3 AAY79584	EGFP sign
9	1258	98.7	239	4 AAB50804	Jellyfish
10	1258	98.7	239	4 AAB85900	A. victor
11	1258	98.7	239	4 AAB31171	Amino aci
12	1258	98.7	239	5 AAG66198	A. victor
13	1258	98.7	239	5 ABG94444	Protease
14	1258	98.7	239	5 AAE14599	Aequorea
15	1258	98.7	239	6 AAG34958	Aequorea
16	1258	98.7	239	6 AAG79829	Green flu
17	1258	98.7	239	6 ABR83616	Green flu
18	1258	98.7	239	6 ADA38074	Aequorea
19	1258	98.7	239	7 ABU63204	Aequorea
20	1258	98.7	239	7 ADC18358	EGFP (enh
21	1258	98.7	239	7 ABW00914	Aequorea
22	1258	98.7	239	7 ADE28570	Enhanced
23	1258	98.7	246	7 ABM79011	Enhanced
24	1258	98.7	248	5 AAG68319	Jellyfish
25	1258	98.7	259	5 AAU99804	Biomembra

## ALIGNMENTS

## RESULT 1

AAE17518  
ID AAE17518 standard; protein; 239 AA.

XX AC AAE17518;

XX AC AAE17518;

DT 22-APR-2002 (first entry)

XX DE Enhanced F64L-E222G jellyfish green fluorescent protein mutant.

XX JW Jellyfish; green fluorescent protein; GFP; protein redistribution;

KW cellular function; genetic reporter; mutant; Stoke's shift; muten.

XX OS Aequorea victoria.

OS Synthetic.

XX FH Key

FT Misc-difference 65

FT Key

FT Misc-difference 65

FT Key

FT Misc-difference 223

FT Key

FT Misc-difference 223

FT Key

FT Misc-difference 223

FT Key

FT Misc-difference 223

FT Key

FT Misc-difference 223

FT Key

FT Misc-difference 223

FT Key

FT Misc-difference 223

FT Key

FT Misc-difference 223

FT Key

FT Misc-difference 223

FT Key

FT Misc-difference 223

FT Key

FT Misc-difference 223

FT Key

FT Misc-difference 223

FT Key

FT Misc-difference 223

FT Key

FT Misc-difference 223

FT Key

FT Misc-difference 223

FT Key

FT Misc-difference 223

FT Key

FT Misc-difference 223

FT Key

XX The invention relates to a fluorescent protein derived from green  
 CC fluorescent protein (GFP) or its analogue. The GFP containing mutations  
 CC at F64L and E222G has a bigger compared to other GFP's making it very  
 CC suitable for high throughput screening due to better resolution. The  
 CC fluorescent protein is useful in *in vitro* assays for measuring protein  
 CC kinase activity or dephosphorylation activity, or for measuring protein  
 CC redistribution. The fluorescent protein is useful in studying cellular  
 CC functions in living cells; as protein tags in transgenic animals, living  
 CC and fixed cells; organelle tags, secretion marker and genetic reporter.  
 CC The fluorescent protein is also useful as a cell or organelle integrity  
 CC marker, a marker for changes in cell morphology, as transfection marker,  
 CC and as a marker to be used in combination with fluorescence activated  
 CC cell sorting (FACS). The novel proteins can also be used as reporters to  
 CC monitor live or dead biomass of organisms, such as fungi. The fluorescent  
 CC protein is also useful as markers in transcriptional and translational  
 CC fusions for performing transposon vector mutagenesis and as a reporter  
 CC for bacterial detection. Transposons encoding the fluorescent protein are  
 CC useful for screening promoters and for tagging plasmids and chromosomes.  
 CC The fluorescent protein engineered into the genome of a phage is useful  
 CC for designing diagnostic tool. The present sequence is a DNA encoding  
 CC enhanced F64L-E222G jellyfish green fluorescent protein (GFP) mutant  
 XX  
 SQ Sequence 239 AA;

Query Match 99.6%; Score 1269; DB 5; Length 239;  
 Best Local Similarity 99.8%; Pred. No. 4.6e-122;  
 Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFVSVSSEGDATYVKLTLPFICTTGKLPVPWPT 60  
 DB 1 MVSKEELFTGVVPIILVELDGDVNGHKFVSVSSEGDATYVKLTLPFICTTGKLPVPWPT 60

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVIIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVIIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239  
 DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239

RESULT 2  
 ABR40352  
 ID ABR40352 standard; protein; 363 AA.  
 XX  
 AC ABR40352;  
 XX  
 DT 08-JUL-2003 (first entry)  
 XX  
 DE Human amino acid sequence SEQ ID NO: 6.  
 XX  
 DE Human; heterologous conjugate; intracellular protein.  
 KW  
 XX Homo sapiens.  
 OS Aequoria victoria.  
 XX  
 XX WO2003029827-A2.  
 PN  
 XX 10-APR-2003.  
 PD  
 XX 01-OCT-2002; 2002WO-DK000651.  
 XX  
 PF 01-OCT-2001; 2001DK-00001433.  
 PR 11-OCT-2001; 2001US-032896P.  
 XX  
 XX (BIOL-) BIOLMAGE AS.  
 PA  
 XX Terry BR, Nielsen SJ;  
 PI

XX WPI: 2003-430211/40.  
 DR N-PSDB; ACC72604.  
 XX  
 PT Novel cell for identifying modulators of protein interaction, contains a  
 PT first conjugate comprising anchor protein, second conjugate having type B  
 PT interactor protein and third conjugate with detectable group.  
 XX  
 PS Disclosure; Page 112-113; 118pp; English.  
 XX  
 CC The invention relates to a novel cell, comprising three heterologous  
 CC conjugates (HC), a first HC (HC1) comprising an anchor protein that  
 CC specifically binds to an internal structure within the cell conjugated to  
 CC an interactor protein (IP) of type A, a second HC (HC2) comprising IP of  
 CC type B conjugated to a first protein of interest, and a third HC (HC3)  
 CC comprising a second protein of interest conjugated to detectable group.  
 CC The cell is useful for detecting if a compound disrupts or induces the  
 CC interaction between two intracellular proteins. The cell is also useful  
 CC for screening compounds that modulate the interaction between two  
 CC intracellular proteins. The present sequence is used in the  
 CC exemplification of the invention  
 XX  
 SQ Sequence 363 AA;

Query Match 99.6%; Score 1269; DB 6; Length 363;  
 Best Local Similarity 99.6%; Pred. No. 8.6e-122;  
 Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFVSVSSEGDATYVKLTLPFICTTGKLPVPWPT 60  
 DB 1 MVSKEELFTGVVPIILVELDGDVNGHKFVSVSSEGDATYVKLTLPFICTTGKLPVPWPT 60

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVIIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVIIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239  
 DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239

RESULT 3  
 AAG65781  
 ID AAG65781 standard; protein; 893 AA.  
 XX  
 AC AAG65781;  
 XX  
 DT 07-JAN-2002 (first entry)  
 XX  
 DE Amino acid sequence of HSPDE4A1-E222G fusion protein.  
 XX  
 DE PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;  
 KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;  
 KW fusion protein.  
 XX  
 XX Homo sapiens.  
 OS Aequoria victoria.  
 XX  
 XX WO200179526-A2.  
 PN  
 XX 25-OCT-2001.  
 PD  
 XX 11-APR-2001; 2001WO-DK000264.  
 XX  
 PF 17-APR-2000; 2000DK-00000651.  
 PR 23-MAY-2000; 2000DK-00000849.  
 XX  
 XX (BIOI-) BIOIMAGE AS.  
 PA  
 XX Terry BR, Nielsen SJ;  
 PI

PI Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;  
PI Praestegaard M;  
DR WPI; 2001-611727/70.  
DR N-PSDB; AAI66852.  
XX  
PT Determining if a compound is a dislocator of PDE4 for identifying  
PT compounds for treating CNS and inflammatory disease comprises identifying  
PT compounds which remove PDE4 spots.  
XX  
PS Example 1; Page 156-160; 160pp; English.  
XX  
CC The invention relates to determining, if a compound, is a dislocator of  
CC PDE4. The method comprises testing if the compound removes PDE4 spots,  
CC which may optionally be induced by a Rolipram-like reference compound,  
CC and testing if it inhibits the catalytic activity of the PDE4, where the  
CC compound is a dislocator of PDE4, if it removes PDE spots and if it does  
CC not inhibit the catalytic activity of PDE4. The method is useful for  
CC identifying compounds useful for the treatment of diseases of the central  
CC nervous system such as depression and for the treatment of inflammatory  
CC disease such as joint inflammation, Crohn's disease, inflammatory bowel  
CC disease, respiratory diseases, chronic obstructive pulmonary disease  
CC (COPD), including asthma, chronic bronchitis, pulmonary emphysema,  
CC endotoxic shock, toxic shock syndrome, systemic lupus erythematosus,  
CC psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV  
CC infection. The use of a reagent that can mimic or reverse the effect of  
CC the compound with affinity for the catalytic site on intracellular  
CC distribution of the PDE for the preparation of a medicament. The present  
CC sequence represents the amino acid sequence of a HSPDE4A1-E222G fusion  
CC protein  
XX  
SQ Sequence 893 AA;  
  
Query Match 99.6%; Score 1269; DB 4; Length 893;  
Best Local Similarity 99.6%; Pred. No. 3.2e-121;  
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 MYSKGEEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGKLTFLKFICTTGKLPVPWPT 60  
Db 655 MYSKGEEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGKLTFLKFICTTGKLPVPWPT 714  
  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 715 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 774  
  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
Db 775 VNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVQLA 834  
  
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVAGFVTAAGITLGMDELYK 239  
Db 835 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVAGFVTAAGITLGMDELYK 893  
  
RESULT 4  
AAG65782  
ID AAG65782 standard; protein; 1132 AA.  
XX  
AC AAG65782;  
XX  
DT 07-JAN-2002 (first entry)  
XX  
DE Amino acid sequence of HSPDE4A4-E222G fusion protein.  
XX  
KW PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;  
KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;  
KW fusion protein.  
XX  
OS Homo sapiens.  
OS Aequorea victoria.  
XX  
PN WO200179526-A2.  
XX

PD 25-OCT-2001.  
XX  
PF 11-APR-2001; 2001WO-DK000264.  
XX  
PR 17-APR-2000; 2000DK-00000651.  
PR 29-MAY-2000; 2000DK-00000849.  
XX  
PA (BIOI-) BIOIMAGE AS.  
XX  
PI Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;  
PI Praestegaard M;  
XX  
DR WPI; 2001-611727/70.  
DR N-PSDB; AAI66853.  
XX  
PT Determining if a compound is a dislocator of PDE4 for identifying  
PT compounds for treating CNS and inflammatory disease comprises identifying  
PT compounds which remove PDE4 spots.  
XX  
PS Example 1; Page 162-167; 160pp; English.  
XX  
CC The invention relates to determining, if a compound, is a dislocator of  
CC PDE4. The method comprises testing if the compound removes PDE4 spots,  
CC which may optionally be induced by a Rolipram-like reference compound,  
CC and testing if it inhibits the catalytic activity of the PDE4, where the  
CC compound is a dislocator of PDE4, if it removes PDE spots and if it does  
CC not inhibit the catalytic activity of PDE4. The method is useful for  
CC identifying compounds useful for the treatment of diseases of the central  
CC nervous system such as depression and for the treatment of inflammatory  
CC disease such as joint inflammation, Crohn's disease, inflammatory bowel  
CC disease, respiratory diseases, chronic obstructive pulmonary disease  
CC (COPD), including asthma, chronic bronchitis, pulmonary emphysema,  
CC endotoxic shock, toxic shock syndrome, systemic lupus erythematosus,  
CC psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV  
CC infection. The use of a reagent that can mimic or reverse the effect of  
CC the compound with affinity for the catalytic site on intracellular  
CC distribution of the PDE for the preparation of a medicament. The present  
CC sequence represents the amino acid sequence of a HSPDE4A4-E222G fusion  
CC protein  
XX  
SQ Sequence 1132 AA;  
  
Query Match 99.6%; Score 1269; DB 4; Length 1132;  
Best Local Similarity 99.6%; Pred. No. 4.5e-121;  
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 MYSKGEEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGKLTFLKFICTTGKLPVPWPT 60  
Db 894 MYSKGEEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGKLTFLKFICTTGKLPVPWPT 953  
  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 954 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 1013  
  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
Db 1014 VNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVQLA 1073  
  
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVAGFVTAAGITLGMDELYK 239  
Db 1074 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVAGFVTAAGITLGMDELYK 1132  
  
RESULT 5  
AAE17517  
ID AAE17517 standard; protein; 239 AA.  
XX  
AC AAE17517;  
XX  
DT 22-APR-2002 (first entry)  
XX  
DE Enhanced F64L jellyfish green fluorescent protein mutant.  
XX

KW Jellyfish; green fluorescent protein; GFP; protein redistribution;  
 KW cellular function; genetic reporter; mutant; Stoke's shift; muten.  
 XX  
 OS Aequorea victoria.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 65 /notes="Wild type Phe substituted with Leu; This  
 FT corresponds to position 64 in the wild type protein"  
 XX  
 FT WO200198338-A2.  
 PN  
 XX  
 XX 27-DEC-2001.  
 XX  
 XX 19-JUN-2001; 2001WO-EP006848.  
 PF  
 XX 19-JUN-2000; 2000DK-00000953.  
 PR 20-JUN-2000; 2000US-0212681P.  
 PR 10-MAY-2001; 2001DK-00000739.  
 PR 10-MAY-2001; 2001US-0290170P.  
 XX  
 XX (BIOI-) BIOIMAGE AS.  
 PA  
 XX Bjorn SP, Pagliaro L, Thastrup O;  
 PI  
 XX WPI; 2002-098224/13.  
 DR N-PSDB; AAD28162.  
 DR  
 XX Novel fluorescent protein in in vitro assay for measuring protein kinase  
 PT activity or dephosphorylation activity, or for measuring protein  
 PT redistribution, has a green fluorescent protein with F64L and E222G  
 PT mutation.  
 XX  
 XX Example 1; Page 35; 41pp; English.  
 PS  
 XX The invention relates to a fluorescent protein derived from green  
 XX fluorescent protein (GFP) or its analogue. The GFP containing mutations  
 CC at F64L and E222G has a bigger compared to other GFP's making it very  
 CC suitable for high throughput screening due to better resolution. The  
 CC fluorescent protein is useful in invitro assays for measuring protein  
 CC kinase activity or dephosphorylation activity, or for measuring protein  
 CC distribution. The fluorescent protein is useful in studying cellular  
 CC functions in living cells; as protein tags in transgenic animals, living  
 CC and fixed cells; organelle tags, secretion marker and genetic reporter.  
 CC The fluorescent protein is also useful as a cell or organelle integrity  
 CC marker, a marker for changes in cell morphology, as transfection marker,  
 CC and as a marker to be used in combination with fluorescence activated  
 CC cell sorting (FACS). The novel proteins can also be used as reporters to  
 CC monitor live or dead biomass of organisms, such as fungi. The fluorescent  
 CC protein is also useful as markers in transcriptional and translational  
 CC fusions for performing transposon vector mutagenesis and as a reporter  
 CC for bacterial detection. Transposons encoding the fluorescent protein are  
 CC useful for screening promoters and for tagging plasmids and chromosomes.  
 CC The fluorescent protein engineered into the genome of a phage is useful  
 CC for designing diagnostic tool. The present sequence is enhanced F64L  
 CC Jellyfish green fluorescent protein (GFP) mutant  
 XX  
 XX Sequence 239 AA;  
 SQ  
 Query Match 99.0%; Score 1261; DB 5; Length 239;  
 Best Local Similarity 99.2%; Pred.No.3.1e-121;  
 Matches :37; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MWSKGEELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTFLKFICTTKGLPVPWPT 60  
 DB 1 MWSKGEELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTFLKFICTTKGLPVPWPT 60  
 QY 51 LVTTLSYGVCFSRYPDHMKQHDFFKSAKMPGVGVQERTIFFKDDGNKTRAEVKEGDTL 120  
 DB 51 LVTTLSYGVCFSRYPDHMKQHDFFKSAKMPGVGVQERTIFFKDDGNKTRAEVKEGDTL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIIMADKQKNGIKVNFKIRHNIEDGVSQLA 180

DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIIMADKQKNGIKVNFKIRHNIEDGVSQLA 180  
 QY 181 DHYQQNTPIGDGFVLLPDNHYLSTQSALSKDPNEKRDHVMVLAVFTAAAGITIGMDLYK 239  
 DB 181 DHYQQNTPIGDGFVLLPDNHYLSTQSALSKDPNEKRDHVMVLAVFTAAAGITIGMDLYK 239  
 RESULT 6  
 ID AAB22882  
 XX AAB22882 standard; protein; 239 AA.  
 AC AAB22882;  
 XX  
 DT 10-JAN-2001 (first entry)  
 XX  
 DE Enhanced green fluorescent protein (EGFP), SEQ ID NO:46.  
 XX  
 KW Bioreactor protein; fusion protein; recognition site;  
 KW cellular targeting sequence; cellular localisation; fluorescent protein;  
 KW protease activity detection; toxin detection; cellular stress detection;  
 KW drug discovery; cell based screening.  
 XX  
 OS Aequorea victoria.  
 OS Synthetic.  
 XX  
 PN WO200050872-A2.  
 XX  
 XX 31-AUG-2000.  
 XX  
 PF 25-FEB-2000; 2000WO-US004794.  
 XX  
 PR 26-FEB-1999; 99US-0122152P.  
 PR 08-MAR-1999; 99US-0123399P.  
 PR 12-JUL-1999; 99US-00352171.  
 XX  
 XX (CELL-) CELLOMICS INC.  
 PA  
 XX Giuliano KA, Kapur R;  
 PI  
 XX WPI; 2000-594086/56.  
 DR N-PSDB; AAA93373.  
 DR  
 XX Automated cell-based characterization of toxin by contacting cells  
 PT containing luminescent reporter molecules with test substance and  
 PT analyzing optically.  
 XX  
 XX Example 11; Fig 29A; 336pp; English.  
 CC The invention relates to systems, methods and reagents for cell-based  
 CC screening or detection of compounds which affect particular biological  
 CC functions. The methods of the invention utilise fluorescent bioreactor  
 CC molecules which, when acted on by a compound of interest, cause an  
 CC alteration in the cellular distribution of at least the fluorescent  
 CC moiety. In one embodiment, the biosensors comprise heat shock proteins  
 CC (HSPs) fused to a fluorescent protein (e.g., jellyfish green fluorescent  
 CC protein (GFP), or derivatives thereof). Such biosensors are located in  
 CC the cytoplasm, but on stress activation translocate to the nucleus. In  
 CC another embodiment bioreactor proteins can be used to detect protease  
 CC activity. Such protease bioreactor fusion proteins comprise one or more  
 CC fluorescent proteins; a recognition signal which is cleaved by the  
 CC protease; and at least one cellular localisation signal. The latter two  
 CC components may be components of a single protein which is acted upon by  
 CC the protease, or may be from heterologous sources. Due to the  
 CC localisation signal, the bioreactor protein is localised to a particular  
 CC region of the cell. Once acted on by the protease of interest, the  
 CC fluorescent protein is cleaved from the localisation sequence, and is  
 CC free to migrate to other locations within the cell. The presence of a  
 CC second localisation signal attached to the fluorescent protein enables  
 CC the fluorescent protein to be directed to a different cellular  
 CC compartment after cleavage of the protease recognition sequence. The  
 CC change in distribution of the fluorescent protein can be detected using  
 CC imaging methods with a high degree of spatial resolution. The methods and



CC biosensors of the invention can be used to investigate a wide range of  
CC cellular activities and to screen compounds which modulate these  
CC activities. Biosensors containing a recognition site for caspase, for  
CC example, may be used for the screening of compounds which modulate  
CC apoptosis, while biosensors containing other protease recognition sites  
CC may be used for the detection of proteolytic toxins (such as anthrax  
CC lethal factor). The method provides improved target validation and  
CC candidate compound optimisation by combining many cell screening formats  
CC with fluorescence-based molecular reagents and computer-based feature  
CC extraction, data analysis and automation, resulting in increased quantity  
CC and speed of data collection and faster evaluation of drug candidates.  
CC Sequences AAB22881-B22885 represent fluorescent proteins which may be used  
CC as components of biosensor fusion proteins of the invention  
XX  
SQ Sequence 239 AA;

Query Match 98.7%; Score 1258; DB 3; Length 239;  
Best Local Similarity 98.7%; Pred. No. 6.3e-121;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTILFKICTTIGKLPVWPWT 60  
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTILFKICTTIGKLPVWPWT 60  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLAFVTAAGITILGMDELYK 239  
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLAFVTAAGITILGMDELYK 239

RESULT 7  
AA54349  
ID AAY54349 standard; protein; 239 AA.  
XX AAY54349;  
XX AC AAY54349;  
XX XX  
XX 06-APR-2000 (first entry)  
XX Amino acid sequence of the mutant green fluorescent protein EGFP.  
DE Fluorescent protein; green fluorescent protein; emission intensity;  
XX fluorescence; pH detection; pH sensor; EGFP.  
KW Synthetic.  
XX Aequorea victoria.  
OS  
FH Key Location/Qualifiers  
FT Misc-difference 65  
FT /note= "wild type Phe substituted with Leu"  
FT Misc-difference 66  
FT /note= "wild type Ser substituted with Thr"  
FT Misc-difference 232  
FT /note= "wild type His substituted with Leu"  
XX  
XX WO964592-A2.  
XX  
XX 16-DEC-1999.  
PD  
XX 08-JUN-1999; 99WO-US012850.  
XX  
XX 09-JUN-1998; 98US-00094359.  
XX 13-OCT-1998; 98US-00172063.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
XX (UYOR-) UNIV OREGON STATE.  
XX

PI Tsien RY, Llopis J, Wachter RM;  
XX WPI: 2000-116540/10.  
DR N-PSDB; AA245642.  
XX  
XX New functional engineered green fluorescent proteins, used for measuring  
PT the pH in biological samples and cells.  
XX  
XX Disclosure; Page 9; 89pp; English.  
XX  
XX The present sequence represents a functional engineered fluorescent  
CC protein based on the Aequorea green fluorescent protein (GFP). The  
CC emission intensity changes as pH varies between 5 and 10 of the present  
CC protein are novel. The functional engineered fluorescent proteins show  
CC reversible changes in fluorescence over physiological pH ranges. They can  
CC be used for determining the pH of samples and cells. The polynucleotides  
CC can also be used to produce transgenic animals. The fluorescent protein  
CC pH sensors can be delivered to cells in the form of polynucleotides  
CC encoding the protein sensor fused to a targeting signal. The targeting  
CC signal directs the expression of the protein sensors to restricted cell  
CC locations. This makes it possible to measure the pH of a precisely  
CC defined cellular region or organelle  
XX  
SQ Sequence 239 AA;

Query Match 98.7%; Score 1258; DB 3; Length 239;  
Best Local Similarity 98.7%; Pred. No. 6.3e-121;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTILFKICTTIGKLPVWPWT 60  
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTILFKICTTIGKLPVWPWT 60  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLAFVTAAGITILGMDELYK 239  
DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLAFVTAAGITILGMDELYK 239

RESULT 8  
AA54349  
ID AAY79584 standard; peptide; 239 AA.  
XX AAY79584;  
XX AC AAY79584;  
XX XX  
XX 29-AUG-2000 (first entry)  
XX EGFP signal domain.  
XX  
XX Protease; biosensor; EGFP; signal peptide; cell screening; assay;  
KW analysis; drug discovery.  
XX  
XX Unidentified.  
XX  
XX WO200026408-A2.  
XX  
XX 11-MAY-2000.  
PD  
XX 29-OCT-1999; 99WO-US025431.  
XX  
XX 30-OCT-1998; 98US-0106308P.  
XX 26-MAY-1999; 99US-0136078P.  
XX  
XX (CELL-) CELLOMICS INC.  
XX  
XX Guiliano KA, Bright G, Olson K, Burroughs-Tencza S;  
PI

XX. WPI; 2000-365644/31.  
DR N-PSDB; AAA27573.  
XX Recombinant nucleic acid encoding a protease biosensor useful for  
PT fluorescence based cell and molecular biochemical assays for drug  
PT discovery comprising three operably linked nucleic acid sequences.  
XX  
XX Claim 14; Fig 29A; 218pp; English.  
XX  
XX The present sequence is that of the EGFP signal domain, which can be  
CC included in novel recombinant protease biosensors (PBs) of the invention.  
CC The PBs (see AAY79638-54) comprise: a first domain (see AAY79579-87)  
CC comprising at least 1 detectable polypeptide signal such as the present  
CC sequence; a second domain (see AAY79588-622) comprising at least 1  
CC protease recognition site; and a third domain (see AAY79623-37)  
CC comprising at least 1 reactant target sequence. A recombinant nucleic  
CC acid (see AAA27627-43) encoding the PB, an expression vector, and a  
CC genetically engineered host cell are also claimed. A claimed method for  
CC identifying compounds that modify protease activity in a cell involves  
CC contacting a host cell that possesses the recombinant PB with a test  
CC compound, and determining the PB distribution in the host cell, where  
CC changes in the distribution of the PB are correlated with modification of  
CC protease activity by the test compound. Claimed kits for identifying  
CC compounds that modify protease activity in a host cell include the  
CC recombinant nucleic acid, or the recombinant PB, or the vector, or the  
CC host cell. The PB is useful in high content screens to detect in vivo  
CC activation of enzymatic activity, and to identify specific activity based  
CC on cleavage of a known recognition motif  
XX  
SQ Sequence 239 AA;

Query Match 98.7%; Score 1258; DB 3; Length 239;  
Best Local Similarity 98.7%; Pred. No. 6.3e-121;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLPFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLPFICTTGKLPVPWPT 60  
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVOLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVOLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLGFVTAAGITLGMDELYK 239

RESULT 9  
AAB50804  
ID AAB50804 standard; protein; 239 AA.  
XX  
AC AAB50804;  
XX  
DT 14-MAR-2001 (first entry)  
XX  
DE Jellyfish GFP mutant EGFP.  
XX  
KW Aequorea victoria; jellyfish; fluorescent protein indicator;  
KW green fluorescent protein; GFP; linker moiety; sensor;  
KW calmodulin-binding domain; mutant; mutein.  
XX  
OS Aequorea victoria.  
XX  
XX WO200071565-A2.  
XX  
XX 30-NOV-2000.  
XX

PF 17-MAY-2000; 2000WO-US013684.  
XX  
XX 21-MAY-1999; 99US-00316919.  
PR 21-MAY-1999; 99US-00316920.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
XX  
XX Tsien RY, Baird GA;  
XX  
XX WPI; 2001-032017/04.  
DR N-PSDB; AAC90488.  
XX  
XX Novel fluorescent proteins comprising a sensor protein inserted into  
PT them, useful for measuring the response of a sensor biological, chemical,  
PT electrical or physiological parameter in vivo or in vitro.  
XX  
XX Disclosure; Page 24; 94pp; English.  
XX  
XX The present sequence is a fluorescent protein used in the construction of  
CC a fluorescent protein indicator. The indicator comprises a sensor  
CC polypeptide that is responsive to a chemical, biological, electrical or  
CC physiological parameter, and a fluorescence protein functional group. The  
CC sensor polypeptide is operatively inserted into the fluorescent moiety.  
CC The fluorescent indicator is useful for detecting the presence of a  
CC response inducing member in a sample. The method involves contacting the  
CC sample with the indicator and detecting a change in fluorescence, in  
CC which a change is indicative of the effect of the parameter on the sensor  
CC polypeptide. The novel fluorescent proteins are advantageous due to their  
CC reduced size as compared to the FRET (fluorescence resonance energy  
CC transfer)-based sensors  
XX  
SQ Sequence 239 AA;

Query Match 98.7%; Score 1258; DB 4; Length 239;  
Best Local Similarity 98.7%; Pred. No. 6.3e-121;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLPFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLPFICTTGKLPVPWPT 60  
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVOLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVOLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLGFVTAAGITLGMDELYK 239

RESULT 10  
AAB85900  
ID AAB85900 standard; protein; 239 AA.  
XX  
AC AAB85900;  
XX  
DT 30-NOV-2001 (first entry)  
XX  
DE A. victoria green fluorescent protein (GFP) and linker sequence.  
XX  
KW Melanin concentrating hormone receptor; MCHR; MCH; chimeric; fusion;  
KW fluorescent polypeptide; orexigenic; anabolic; food intake; GFP;  
KW green fluorescent protein.  
XX  
OS Synthetic.  
OS Aequorea victoria.  
XX  
XX WO200168706-A1.  
XX

PD 20-SEP-2001.  
 XX 14-MAR-2001; 2001WO-US008071.  
 PF  
 XX 15-MAR-2000; 2000US-0189698P.  
 PR  
 XX (MERI ) MERCK & CO INC.  
 PA  
 XX Marsh DJ;  
 XX  
 XX WPI; 2001-565791/63.  
 DR N-PSDB; AAH47304.  
 DR  
 XX Fusion proteins comprising melanin concentrating hormone receptor  
 PT peptides and fluorescent proteins, useful for identifying appetite  
 PT stimulants.  
 PT  
 XX Claim 2; Page 14; 71pp; English.  
 PS  
 XX The invention provides melanin concentrating hormone (MCH) receptor  
 CC (MCHR) chimeric and fusion proteins. The MCHR chimeric proteins comprise  
 CC MCHR polypeptide regions from different species. The MCHR fusion protein  
 CC comprise MCHR polypeptide region and a fluorescent polypeptide region  
 CC joined directly, or via a linker, to the carboxy side of the MCHR  
 CC polypeptide region. The MCHR fusion proteins can be expressed by standard  
 CC recombinant methodology. MCH action promotes feeding (orexigenic) and up  
 CC regulation of MCH activity stimulates food intake. The present sequence  
 CC represents a A. victoria green fluorescent protein (GFP) and a linker  
 CC sequence  
 CC  
 XX Sequence 239 AA;  
 SQ

Query Match 98.7%; Score 1258; DB 4; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 6.3e-121; Indels 0; Gaps 0;  
 Matches 236; Conservative 1; Mismatches 2;  
 QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60  
 DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60  
 QY 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGSGVQLA 180  
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGSGVQLA 180  
 QY 181 DHYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMLAGFVTAAGITLGMDELYK 239  
 DB 181 DHYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMLVLEFVTAAGITLGMDELYK 239

RESULT 11  
 AAB31171  
 ID AAB31171 standard; protein; 239 AA.  
 XX  
 AC AAB31171;  
 XX  
 XX 02-APR-2001 (first entry)  
 DT  
 XX Amino acid sequence of a green fluorescent protein (GFP).  
 DE  
 XX Growth rate; death rate; reporter gene; luminescent protein;  
 KW fluorescent product; luciferase; green fluorescent protein; GFP.  
 XX  
 XX Aequorea victoria.  
 OS  
 XX WO200075367-A1.  
 PN  
 XX 14-DEC-2000.  
 PD  
 XX 07-JUN-2000; 2000WO-FI000507.  
 PF

XX 07-JUN-1999; 99FI-00001296.  
 PR (LILI/) LILIUS E.  
 PA (VIRT/) VIRT M.  
 XX  
 PI Lilius E, Virta M;  
 XX WPI; 2001-061737/07.  
 DR N-PSDB; AAC86954.  
 DR  
 XX Assessing growth and death rates of a micro-organism in a desired  
 PT environment, by introducing 2 reporter genes encoding luminescent and  
 PT fluorescent products and detecting luminescent fluorescence.  
 XX  
 PS Disclosure; Page 27; 32pp; English.  
 PS  
 XX The specification describes a method for assessing the growth rate and  
 CC death rate of a micro-organism within a predetermined time period in a  
 CC desired environment. The method comprises introducing at least two  
 CC reporter genes encoding luminescent and/or fluorescent products into the  
 CC micro-organisms, incubating the micro-organism within the desired  
 CC environment, and detecting luminescence and/or fluorescence after a  
 CC predetermined time period. Use of two different markers within a micro-  
 CC organism enables the differentiation between growth and death rates. The  
 CC method is used to assess the growth rate and death rate of a micro-  
 CC organism within a predetermined time period in a desired environment. The  
 CC present sequence represents a green fluorescent protein (GFP), and is  
 CC encoded by a plasmid which encodes luminescent and fluorescent proteins,  
 CC and is used in the method of the invention  
 CC  
 XX Sequence 239 AA;  
 SQ

Query Match 98.7%; Score 1258; DB 4; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 6.3e-121; Indels 0; Gaps 0;  
 Matches 236; Conservative 1; Mismatches 2;  
 QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60  
 DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60  
 QY 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGSGVQLA 180  
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGSGVQLA 180  
 QY 181 DHYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMLAGFVTAAGITLGMDELYK 239  
 DB 181 DHYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMLVLEFVTAAGITLGMDELYK 239

RESULT 12  
 AAG66198  
 ID AAG66198 standard; protein; 239 AA.  
 XX  
 AC AAG66198;  
 XX  
 XX 17-JUN-2002 (first entry)  
 DT  
 XX A. victoria green fluorescent protein (EGFP).  
 DE  
 XX Cyan-green fluorescent protein; fluorescence; recombinant; GFP;  
 KW green fluorescent protein; EGFP.  
 XX  
 XX Aequorea victoria.  
 OS  
 XX JP20002045189-A.  
 PN  
 XX 12-FEB-2002.  
 PD  
 XX

```

PF 04-AUG-2000; 2000JP-00237165.
XX
PR 04-AUG-2000; 2000JP-00237165.
XX
XX (RIKA ) RIKAGAKU KENKYUSHO.
XX
XX WPI; 2002-299190/34.
XX
XX N-PSDB; ABL40628.
XX
XX A gene encoding cyan-green fluorescent protein.
XX
XX Example; Page 14; 20pp; Japanese.
XX
XX The invention relates to a gene encoding proteins having cyan-green
XX fluorescence characteristic and having a function of showing stable
XX fluorescence characteristic in acid region. A method for the preparation
XX of a cyan-green fluorescent protein is provided which involves a
XX transformant transformed by a recombinant vector comprising the gene,
XX where the transformant is cultured and the protein is collected from the
XX culture. The present sequence represents the A. victoria green
XX fluorescent protein (EGFP)
XX
XX Sequence 239 AA;
XX
XX Query Match 98.7%; Score 1258; DB 5; Length 239;
XX Best Local Similarity 98.7%; Pred. No. 6.3e-121;
XX Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 MVSKEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGKLTLLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGKLTLLKFICTTGKLPVPWPT 60
QY 61 LVTTLTSGVQCFSRYDPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTSGVQCFSRYDPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSGVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSGVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239
RESULT 13
ABG9444
ID ABG94444 standard; protein; 239 AA.
AC ABG94444;
XX
XX 27-NOV-2002 (first entry)
XX
XX Protease biosensor signal sequence #6.
XX
XX Detection; classification; identification; toxin detection; protease;
XX ADP-ribosylating toxin; cytotoxic phospholipase; exfoliative toxin;
XX toxic threat agent.
XX
XX Synthetic.
XX
XX US6416959-B1.
XX
XX 09-JUL-2002.
XX
XX 25-FEE-2000; 2000US-00513783.
XX
XX 27-FEE-1997; 97US-00810983.
XX
XX 27-FEE-1998; 98US-00031271.
XX
XX 26-FEE-1999; 99US-0122152P.
XX
XX 08-MAR-1999; 99US-0123399P.
XX
XX 12-JUL-1999; 99US-00352171.
XX
XX 31-AUG-1999; 99US-0151797P.
XX

```

---

```

PR 17-SEP-1999; 99US-00398965.
PR 29-OCT-1999; 99US-00430656.
PR 01-DEC-1999; 99US-0168408P.
XX
XX (GIUL/) GIULIANO K.
XX (KAPU/) KAPUR R.
XX
XX Giuliano K, Kapur R;
XX
XX WPI; 2002-634730/68.
XX
XX N-PSDB; ABS71491.
XX
XX Automated cell-based toxin detection, classification, and/or
XX identification by treating cells involves use of three classes of
XX luminescent reporter molecules such as detectors, classifiers or
XX identifiers.
XX
XX Example 10; Fig 29A; 214pp; English.
XX
XX The invention describes methods of automated detection, classification
XX and identification comprising treating cells containing luminescent
XX reporter molecules (I) in array of locations with a test substance, where
XX (I) are detectors, classifiers or identifiers, imaging cells in each
XX location to obtain luminescent signals and converting optical information
XX into digital data to interpret presence of toxins in the test substance.
XX The method are useful for detection of toxins chosen from proteases, ADP-
XX ribosylating toxins, cytotoxic phospholipases, and exfoliative toxins.
XX Three classes of cell-based luminescent reporter molecules such as
XX detectors, classifiers and identifiers are described and serve as
XX reporters of toxic threat agents. The first two levels of
XX characterization enable a rapid readout of toxin class without
XX sacrificing the ability to detect many new mutant toxins or dissect
XX several complex mixtures of known toxins. This is the amino acid sequence
XX of a protease biosensor related signal sequence used in the cell-based
XX screening system
XX
XX Sequence 239 AA;
XX
XX Query Match 98.7%; Score 1258; DB 5; Length 239;
XX Best Local Similarity 98.7%; Pred. No. 6.3e-121;
XX Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 MVSKEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGKLTLLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGKLTLLKFICTTGKLPVPWPT 60
QY 61 LVTTLTSGVQCFSRYDPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTSGVQCFSRYDPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSGVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSGVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239
RESULT 14
AAE14599
ID AAE14599 standard; protein; 239 AA.
XX
XX AAE14599;
XX
XX 31-MAY-2002 (first entry)
XX
XX Aequorea victoria enhanced green fluorescent protein.
XX
XX Mutagenesis; enhanced green fluorescent protein; EGFP; mutant; mutein.
XX
XX Aequorea victoria.
XX
XX Synthetic.

```

XX Key Location/Qualifiers  
 FT Misc-difference 1. .3 /note= "Wild-type GFP Met-Ser are replaced with Met-Val-Ser"  
 FT FT  
 FT FT  
 FT FT  
 FT Misc-difference 65 /note= "GFP Phe64 is replaced by Leu"  
 FT FT  
 FT Misc-difference 66 /note= "GFP Ser65 is replaced by Thr"  
 FT FT  
 PN EP1178109-A1.  
 XX  
 XX 06-FEB-2002.  
 XX  
 XX 03-AUG-2001; 2001BP-00306650.  
 XX  
 XX 04-AUG-2000; 2000JP-00237166.  
 PR  
 XX (RIKE ) RIKEN KK.  
 XX  
 XX Miyawaki A, Sawano A;  
 XX WPI; 2002-208112/27.  
 DR N-PSDB; AAD27910.  
 XX  
 XX Method for mutagenesis, e.g. for introducing certain or random mutations at certain sites of the nucleotide sequence, comprises synthesizing a mutated strand and a complementary strand by use of megaprimers.  
 PT  
 PT  
 PT  
 PS Example 1; Page 13-14; 31pp; English.  
 XX  
 XX The invention relates to a method for mutagenesis that comprises synthesizing a mutated strand and a complementary strand by use of megaprimers. The method basically comprises a DNA synthesis in which one or more primers that have a nucleotide sequence containing at least one mutation and a phosphorylated 5'-terminus are annealed to a template DNA and then subjected to an elongation reaction using a thermostable high-fidelity DNA polymerase, after which the phosphorylated 5'-terminus and the elongated terminus are ligated by means of a thermostable DNA ligase to synthesise a circular DNA containing the primers; a digestion in which the step of DNA synthesis is repeated several times to amplify the DNA containing the primers and then, at least DNAs other than the amplified circular DNA are digested into several fragments; and a double-stranded DNA synthesis in which, with the several fragments obtained in the step of digestion as megaprimers, the megaprimers are annealed to the circular DNA synthesised above, followed by an elongation reaction performed using the thermostable high-fidelity DNA polymerase. The method is useful for mutagenesis, particularly for introducing certain mutations at certain sites of a nucleotide sequence, or for introducing random mutations at certain sites of the nucleotide sequence. The present method is simple, speedy, economical and widely applicable. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used for mutagenesis in an exemplification of the invention. The EGFP is derived by altering the green fluorescent protein (GFP) sequence of Aequorea victoria  
 XX  
 SQ Sequence 239 AA;

Query Match 98.7%; Score 1258; DB 5; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 6.3e-121;  
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYGKLTAKFICTTGKLPVPWPT 60  
 DB 1 MVSKEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYGKLTAKFICTTGKLPVPWPT 60  
 QY 61 LVTTLSYGVCFSRYPDHMKQHDFFPKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTTLTLYGVCFSRYPDHMKQHDFFPKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFEDGNILGHKLEYNHNHVIIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 DB 121 VNRIELKGIDFEDGNILGHKLEYNHNHVIIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 QY 181 DHYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239  
 DB 181 DHYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239

QY 181 DHYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239  
 DB 181 DHYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239

RESULT 15  
 AAEE34958  
 ID AAEE34958 standard; protein; 239 AA.  
 XX  
 AC AAEE34958;  
 XX  
 DT 28-MAY-2003 (first entry)  
 DE  
 DE Aequorea victoria enhanced green fluorescent protein (EGFP).  
 XX  
 KW Phosphorylation indicator; fluorescent protein; detection; phosphatase;  
 KW Kinase; enhanced green fluorescent protein; EGFP.  
 XX  
 OS Aequorea victoria.  
 XX  
 PN WO200295058-A2.  
 XX  
 PD 28-NOV-2002.  
 XX  
 PF 24-MAY-2002; 2002WO-US016955.  
 XX  
 PR 24-MAY-2001; 2001US-00865291.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 XX Tsien RV, Ting AV, Zhang J;  
 XX WPI; 2003-148474/14.  
 DR N-PSDB; AAD53428.  
 XX

Novel chimeric phosphorylation indicators, useful for detecting kinase/phosphatase in samples, has donor molecule, phosphorylatable domain, phosphoaminoacid binding domain, and acceptor molecule, in operative linkage.  
 XX  
 PS Disclosure; Col 56-57; 38pp; English.  
 XX  
 XX The present invention relates to chimeric phosphorylation indicators comprising a phosphorylation polypeptide and a fluorescent protein or in operative linkage, a donor molecule, a phosphorylatable domain, a phosphoaminoacid binding domain (PABD) and an acceptor molecule. The phosphorylation indicators of the invention are useful for detecting kinases or phosphatases in a biological sample. They are also useful in high throughput analysis e.g. for detecting a kinase inhibitor or phosphatase inhibitor. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used in the invention  
 XX  
 SQ Sequence 239 AA;

Query Match 98.7%; Score 1258; DB 6; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 6.3e-121;  
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYGKLTAKFICTTGKLPVPWPT 60  
 DB 1 MVSKEELFTGVVPIILVELDGVNGHKFVSVEGEGDATYGKLTAKFICTTGKLPVPWPT 60  
 QY 61 LVTTLSYGVCFSRYPDHMKQHDFFPKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTTLTLYGVCFSRYPDHMKQHDFFPKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFEDGNILGHKLEYNHNHVIIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 DB 121 VNRIELKGIDFEDGNILGHKLEYNHNHVIIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 QY 181 DHYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239  
 DB 181 DHYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239

Db 181 DHYQONTPIGDGPVLLFPDNHYLSTQSALSXPNEKRDHVMVLLFEFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 15:54:13  
Job time : 48.1111 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:55 ; Search time 12.7778 Seconds  
(without alignments)  
965.630 Million cell updates/sec

Title: US-09-887-784-222a

Perfect score: 1274

Sequence: 1 MYSKGEELFTGVVPIVLVD.....VLAFGFTAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*

- 1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep:\*
- 2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep:\*
- 3: /cgn2\_6/prodata/2/iaa/6A\_COMB.pep:\*
- 4: /cgn2\_6/prodata/2/iaa/6B\_COMB.pep:\*
- 5: /cgn2\_6/prodata/2/iaa/ECTUS\_COMB.pep:\*
- 6: /cgn2\_6/prodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1258	98.7	239	3	US-09-172-063-3 Sequence 3, Appli
2	1258	98.7	239	4	US-09-513-783A-46 Sequence 46, Appl
3	1258	98.7	239	4	US-09-316-919-4 Sequence 4, Appli
4	1258	98.7	239	4	US-09-602-641-3 Sequence 3, Appli
5	1258	98.7	239	4	US-09-920-922-2 Sequence 2, Appli
6	1258	98.7	281	3	US-09-062-102-1 Sequence 1, Appli
7	1258	98.7	281	4	US-09-364-946-1 Sequence 1, Appli
8	1258	98.7	294	4	US-09-513-783A-2 Sequence 2, Appli
9	1258	98.7	323	3	US-09-172-063-21 Sequence 21, Appl
10	1258	98.7	323	4	US-09-602-641-21 Sequence 21, Appl
11	1258	98.7	364	3	US-09-085-305-6 Sequence 6, Appli
12	1258	98.7	379	4	US-09-417-197-129 Sequence 129, Appl
13	1258	98.7	434	4	US-09-800-170-48 Sequence 48, Appl
14	1258	98.7	442	4	US-09-417-197-127 Sequence 127, App
15	1258	98.7	459	4	US-09-513-783A-170 Sequence 170, App
16	1258	98.7	544	4	US-09-417-197-113 Sequence 113, App
17	1258	98.7	544	4	US-09-417-197-115 Sequence 115, App
18	1258	98.7	604	4	US-09-417-197-59 Sequence 59, Appl
19	1258	98.7	605	4	US-09-417-197-41 Sequence 41, Appl
20	1258	98.7	606	4	US-09-417-197-65 Sequence 65, Appl
21	1258	98.7	607	4	US-09-417-197-47 Sequence 47, Appl
22	1258	98.7	630	4	US-09-417-197-63 Sequence 63, Appl
23	1258	98.7	631	4	US-09-417-197-39 Sequence 39, Appl
24	1258	98.7	633	4	US-09-417-197-45 Sequence 45, Appl
25	1258	98.7	635	4	US-09-417-197-125 Sequence 125, App
26	1258	98.7	642	2	US-08-818-253-2 Sequence 2, Appli
27	1258	98.7	642	2	US-08-818-253-6 Sequence 6, Appli

28	1258	98.7	642	3	US-08-818-253-2	Sequence 2, Appli
29	1258	98.7	642	3	US-08-818-253-6	Sequence 6, Appli
30	1258	98.7	652	3	US-08-818-253-4	Sequence 4, Appli
31	1258	98.7	652	3	US-08-818-252-4	Sequence 4, Appli
32	1258	98.7	718	4	US-09-417-197-75	Sequence 75, Appl
33	1258	98.7	719	4	US-09-417-197-51	Sequence 51, Appl
34	1258	98.7	726	4	US-09-417-197-71	Sequence 71, Appl
35	1258	98.7	727	4	US-09-417-197-139	Sequence 139, App
36	1258	98.7	783	4	US-09-513-783A-176	Sequence 176, App
37	1258	98.7	797	4	US-09-417-197-141	Sequence 141, App
38	1258	98.7	797	4	US-09-417-197-143	Sequence 143, App
39	1258	98.7	798	4	US-09-417-197-77	Sequence 77, Appl
40	1258	98.7	805	4	US-09-513-783A-178	Sequence 178, App
41	1258	98.7	806	4	US-09-417-197-51	Sequence 53, Appl
42	1258	98.7	836	4	US-09-417-197-61	Sequence 61, Appl
43	1258	98.7	842	4	US-09-417-197-43	Sequence 43, Appl
44	1258	98.7	843	4	US-09-417-197-117	Sequence 117, App
45	1258	98.7	853	4	US-09-417-197-119	Sequence 119, App

ALIGNMENTS

RESULT 1  
US-09-172-063-3  
; Sequence 3, Application US/09172063  
; Patent No. 6150176  
; GENERAL INFORMATION:  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Miyawaki, Atsushi  
; APPLICANT: Llopis, Juan  
; APPLICANT: Wachter, Rebekka M.  
; APPLICANT: Remington, S. James  
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
; FILE REFERENCE: 07257/071001  
; CURRENT APPLICATION NUMBER: US/09/172,063  
; CURRENT FILING DATE: 1998-10-13  
; EARLIER APPLICATION NUMBER: 09/094,359  
; EARLIER FILING DATE: 1998-06-09  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: EGFP  
US-09-172-063-3

Query Match	98.7%	Score 1258;	DB 3;	Length 239;
Best Local Similarity	98.7%	Pred. No. 4e-128;		
Matches 236;	Conservative 1;	Mismatches 2;	Indels 0;	Gaps 0;
QY	1	MYSKGEELFTGVVPIVLVDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVWPWT	60	
Db	1	MYSKGEELFTGVVPIVLVDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVWPWT	60	
QY	61	LVTTLTSGVQCPSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNGYKTRAEVKFEGDTL	120	
Db	61	LVTTLTSGVQCPSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNGYKTRAEVKFEGDTL	120	
QY	121	VNRIELKIDFKEDGNILGHKLEYNHNHVIMADKQNGIKVNFKIPHNIEDGVSQVLA	180	
Db	121	VNRIELKIDFKEDGNILGHKLEYNHNHVIMADKQNGIKVNFKIPHNIEDGVSQVLA	180	
QY	181	DHYQONTPTGDDGVPVLLPNHNYLSTQSALSKDNEKRDHMLVAGFVTAAGITLGMDELYK	239	
Db	181	DHYQONTPTGDDGVPVLLPNHNYLSTQSALSKDNEKRDHMLVAGFVTAAGITLGMDELYK	239	

```

RESULT 2
US-09-513-783A-46
; Sequence 46, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGFP
US-09-513-783A-46

Query Match          98.7%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 4e-128;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVLAGFVTAAGITILGMDLYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVLAGFVTAAGITILGMDLYK 239

RESULT 3
US-09-316-919-4
; Sequence 4, Application US/09316919
; Patent No. 6469154
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: 07257/073001
; CURRENT APPLICATION NUMBER: US/09/316,919
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-316-919-4

Query Match          98.7%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 4e-128;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVLAGFVTAAGITILGMDLYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVLAGFVTAAGITILGMDLYK 239

RESULT 4
US-09-602-641-3
; Sequence 3, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wichter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-602-641-3

Query Match          98.7%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 4e-128;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVLAGFVTAAGITILGMDLYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVLAGFVTAAGITILGMDLYK 239

RESULT 5
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. 6673610
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9

```



```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; US-09-920-922-2

Query Match      98.7%; Score 1258; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 4e-128;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTILKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTILKFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLAGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLAGFVTAAGITLGMDELYK 239

RESULT 6
US-09-062-102-1
; Sequence 1, Application US/09062102
; Patent No. 6130313
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangliang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100
; CURRENT APPLICATION NUMBER: US/09/062,102
; CURRENT FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: US 60/060,855
; EARLIER FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6130313
US-09-062-102-1

Query Match      98.7%; Score 1258; DB 3; Length 281;
Best Local Similarity 98.7%; Pred. No. 5.1e-128;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTILKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTILKFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLAGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLAGFVTAAGITLGMDELYK 239

RESULT 7
US-09-364-946-1
; Sequence 1, Application US/09364946
; Patent No. 6306600
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangliang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100CJP/D2
; CURRENT APPLICATION NUMBER: US/09/364,946
; CURRENT FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: US 09/191,233
; EARLIER FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6306600
US-09-364-946-1

Query Match      98.7%; Score 1258; DB 4; Length 281;
Best Local Similarity 98.7%; Pred. No. 5.1e-128;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTILKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTILKFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLAGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLAGFVTAAGITLGMDELYK 239

RESULT 8
US-09-513-783A-2
; Sequence 2, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kaput, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GFP-DEVD-Annexin II construct
US-09-513-783A-2

Query Match      98.7%; Score 1258; DB 4; Length 294;
Best Local Similarity 98.7%; Pred. No. 5.5e-128;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTILKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTILKFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLAGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLAGFVTAAGITLGMDELYK 239
```

Db 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTLPVWPWT 60  
QY 61 LVTLLSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
Db 121 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLAGFVTAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLAGFVTAAGITLGMDELYK 239

## RESULT 9

US-09-172-063-21  
; Sequence 21, Application US/09172063  
; Patent No. 6150176  
; GENERAL INFORMATION:  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Miyawaki, Atsushi  
; APPLICANT: Llopis, Juan  
; APPLICANT: Wachter, Rebekka M.  
; APPLICANT: Remington, S. James  
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
; MEASURING THE PH OF A BIOLOGICAL SAMPLE  
; FILE REFERENCE: 07257/071001  
; CURRENT APPLICATION NUMBER: US/09/172,063  
; CURRENT FILING DATE: 1998-10-13  
; EARLIER APPLICATION NUMBER: 09/094,359  
; EARLIER FILING DATE: 1998-06-09  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 21  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: GT-EGFP  
US-09-172-063-21

Query Match 98.7%; Score 1258; DB 3; Length 323;  
Best Local Similarity 98.7%; Pred. No. 6.3e-128;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTLPVWPWT 60  
Db 85 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTLPVWPWT 144  
QY 61 LVTLLSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 145 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 204  
QY 121 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
Db 205 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 264  
QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLAGFVTAAGITLGMDELYK 239  
Db 265 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLAGFVTAAGITLGMDELYK 323

## RESULT 10

US-09-602-641-21  
; Sequence 21, Application US/09602641  
; Patent No. 6608189  
; GENERAL INFORMATION:  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Miyawaki, Atsushi  
; APPLICANT: Llopis, Juan

; APPLICANT: Wachter, Rebekka M.  
; APPLICANT: Remington, S. James  
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
; MEASURING THE PH OF A BIOLOGICAL SAMPLE  
; FILE REFERENCE: 07257/071001  
; CURRENT APPLICATION NUMBER: US/09/602,641  
; CURRENT FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 09/172,063  
; PRIOR FILING DATE: 1998-10-13  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 21  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: GT-EGFP  
US-09-602-641-21

Query Match 98.7%; Score 1258; DB 4; Length 323;  
Best Local Similarity 98.7%; Pred. No. 6.3e-128;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTLPVWPWT 60  
Db 85 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTLPVWPWT 144  
QY 61 LVTLLSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 145 LVTLLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 204  
QY 121 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
Db 205 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 264  
QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLAGFVTAAGITLGMDELYK 239  
Db 265 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLAGFVTAAGITLGMDELYK 323

## RESULT 11

US-09-085-305-6  
; Sequence 6, Application US/09085305  
; Patent No. 6191269  
; GENERAL INFORMATION:  
; APPLICANT: Pollock, Allan  
; APPLICANT: Lovett, David H.  
; APPLICANT: Turck, Johanna  
; TITLE OF INVENTION: Selective Induction of Apoptosis in  
; Malignant Cancer Cells by Delivery of N-Terminal  
; Interleukin-1-Alpha Pro-Piece Polypeptide  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Bozicevic & Reed, LLP  
; STREET: 285 Hamilton Ave, Suite 200  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/085,305  
; FILING DATE: 29-MAY-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Francis, Carol L  
REGISTRATION NUMBER: 36,513  
REFERENCE/DOCKET NUMBER: 6510/102US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-327-3400  
TELEFAX: 650-327-3231  
TELEX:  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 364 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-085-305-6

Query Match 98.7%; Score 1258; DB 3; Length 364;  
Best Local Similarity 98.7%; Pred. No. 7.6e-128;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MYSKGELFTGVVPILVELDGVNKGKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60  
DB 126 MYSKGELFTGVVPILVELDGVNKGKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 185

QY 61 LVTTLSYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 186 LVTTLSYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 245

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQVA 180  
DB 246 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQVA 305

QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLAGFTVTAAGITLGMDELYK 239  
DB 306 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLAGFTVTAAGITLGMDELYK 364

RESULT 12  
US-09-417-197-129  
Sequence 129, Application US/09417197  
Patent No. 6518021  
GENERAL INFORMATION:  
APPLICANT: Ole Thastrup, et al.  
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An  
FILE REFERENCE: 3759-0110P  
CURRENT FILING DATE: 1999-10-07  
NUMBER OF SEQ ID NOS: 143  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 129  
LENGTH: 379  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: actin-binding-domain-EGFP fusion  
US-09-417-197-129

Query Match 98.7%; Score 1258; DB 4; Length 379;  
Best Local Similarity 98.7%; Pred. No. 8e-128;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MYSKGELFTGVVPILVELDGVNKGKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60  
DB 141 MYSKGELFTGVVPILVELDGVNKGKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 200

QY 61 LVTTLSYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 201 LVTTLSYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 260

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQVA 180  
DB 261 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQVA 320

QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLAGFTVTAAGITLGMDELYK 239  
DB 321 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLAGFTVTAAGITLGMDELYK 379

RESULT 13  
US-09-800-170-48  
Sequence 48, Application US/09800170  
Patent No. 6481667  
GENERAL INFORMATION:  
APPLICANT: Kinsella, Todd  
TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES  
FILE REFERENCE: A-68614-1/DJB/RMS/RMK  
CURRENT APPLICATION NUMBER: US/09/800,170  
CURRENT FILING DATE: 2001-08-28  
PRIOR APPLICATION NUMBER: US 60/187,130  
PRIOR FILING DATE: 2000-03-06  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 48  
LENGTH: 434  
TYPE: PRT  
ORGANISM: Synechocystis PCC6803  
US-09-800-170-48

Query Match 98.7%; Score 1258; DB 4; Length 434;  
Best Local Similarity 98.7%; Pred. No. 9.9e-128;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MYSKGELFTGVVPILVELDGVNKGKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60  
DB 196 MYSKGELFTGVVPILVELDGVNKGKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 255

QY 61 LVTTLSYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 256 LVTTLSYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 315

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQVA 180  
DB 316 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGVSQVA 375

QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLAGFTVTAAGITLGMDELYK 239  
DB 376 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLAGFTVTAAGITLGMDELYK 434

RESULT 14  
US-09-417-197-127  
Sequence 127, Application US/09417197  
Patent No. 6518021  
GENERAL INFORMATION:  
APPLICANT: Ole Thastrup, et al.  
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An Ir  
FILE REFERENCE: 3759-0110P  
CURRENT APPLICATION NUMBER: US/09/417,197  
CURRENT FILING DATE: 1999-10-07  
NUMBER OF SEQ ID NOS: 143  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 127  
LENGTH: 442  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: EGFP-RhoA fusion  
US-09-417-197-127

Query Match 98.7%; Score 1258; DB 4; Length 442;  
Best Local Similarity 98.7%; Pred. No. 1e-127;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MYSKGELFTGVVPILVELDGVNKGKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60

```
Db 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGGDATYGLTLKFICTTGKLPVPWPT 60
QY 61 LVTTLISGVQCFSRYPDHMKQHDFFPKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLISGVQCFSRYPDHMKQHDFFPKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLAFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLAFVTAAGITLGMDELYK 239
```

## RESULT 15

```
US-09-513-783A-170
; Sequence 170, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 170
; LENGTH: 459
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP27
US-09-513-783A-170
```

```
Query Match 98.7%; Score 1258; DB 4; Length 459;
Best Local Similarity 98.7%; Pred. No. 1.1e-127; Indels 0; Gaps 0;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGGDATYGLTLKFICTTGKLPVPWPT 60
Db 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGGDATYGLTLKFICTTGKLPVPWPT 60
QY 61 LVTTLISGVQCFSRYPDHMKQHDFFPKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLISGVQCFSRYPDHMKQHDFFPKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLAFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLAFVTAAGITLGMDELYK 239
```

Search completed: June 21, 2004, 16:04:05  
Job time: 13.7778 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:54:20 ; Search time 34.7778 Seconds  
(without alignments)  
1940.117 Million cell updates/sec

Title: US-09-887-784-222a

Perfect score: 1274

Sequence: 1 MVSKEELFTGVVPIVLVELD.....VLaGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1269	99.6	239	9	US-09-887-784-4
2	1269	99.6	239	12	US-10-296-953-4
3	1269	99.6	363	14	US-10-270-223-6
4	1269	99.6	893	14	US-10-257-909A-30
5	1269	99.6	1132	14	US-10-257-909A-32
6	1261	99.0	239	9	US-09-887-784-2
7	1261	99.0	239	12	US-10-296-953-2
8	1258	98.7	239	9	US-09-920-922-2
9	1258	98.7	239	9	US-09-999-745-4
10	1258	98.7	239	10	US-09-866-538-4
11	1258	98.7	239	10	US-09-797-496B-2
12	1258	98.7	239	10	US-09-794-308-4
13	1258	98.7	239	10	US-09-865-291-4
14	1258	98.7	239	12	US-10-457-982-3
15	1258	98.7	239	14	US-10-121-258-13

16	1258	98.7	239	14	US-10-221-461-7	Sequence 7, Appli
17	1258	98.7	239	14	US-10-100-957A-46	Sequence 46, Appli
18	1258	98.7	239	14	US-10-177-390-2	Sequence 2, Appli
19	1258	98.7	239	14	US-10-338-411-3	Sequence 3, Appli
20	1258	98.7	239	15	US-10-370-570-4	Sequence 4, Appli
21	1258	98.7	239	15	US-10-389-640-3	Sequence 3, Appli
22	1258	98.7	259	14	US-10-314-861-11	Sequence 11, Appli
23	1258	98.7	281	12	US-09-931-232-1	Sequence 37, Appli
24	1258	98.7	288	14	US-10-314-861-37	Sequence 11, Appli
25	1258	98.7	293	14	US-10-314-861-35	Sequence 35, Appli
26	1258	98.7	294	14	US-10-100-957A-2	Sequence 2, Appli
27	1258	98.7	295	14	US-10-314-861-39	Sequence 39, Appli
28	1258	98.7	299	14	US-10-314-861-33	Sequence 33, Appli
29	1258	98.7	305	14	US-10-314-861-31	Sequence 31, Appli
30	1258	98.7	308	14	US-10-033-717-35	Sequence 35, Appli
31	1258	98.7	311	14	US-10-314-861-29	Sequence 29, Appli
32	1258	98.7	320	14	US-10-338-411-11	Sequence 11, Appli
33	1258	98.7	320	15	US-10-389-640-11	Sequence 11, Appli
34	1258	98.7	323	12	US-10-457-982-21	Sequence 21, Appli
35	1258	98.7	323	14	US-10-338-411-7	Sequence 7, Appli
36	1258	98.7	323	14	US-10-338-411-13	Sequence 13, Appli
37	1258	98.7	323	15	US-10-389-640-7	Sequence 7, Appli
38	1258	98.7	323	15	US-10-389-640-13	Sequence 13, Appli
39	1258	98.7	324	14	US-10-314-861-16	Sequence 16, Appli
40	1258	98.7	345	14	US-10-338-411-5	Sequence 5, Appli
41	1258	98.7	345	15	US-10-389-640-5	Sequence 5, Appli
42	1258	98.7	346	14	US-10-338-411-9	Sequence 9, Appli
43	1258	98.7	346	15	US-10-389-640-9	Sequence 9, Appli
44	1258	98.7	359	14	US-10-033-717-33	Sequence 33, Appli
45	1258	98.7	359	14	US-10-033-717-34	Sequence 34, Appli

ALIGNMENTS

RESULT 1

US-09-887-784-4  
; Sequence 4, Application US/09887784  
; Patent No. US20020177189A1  
; GENERAL INFORMATION:  
; APPLICANT: BJORN, Sara et al  
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS  
; FILE REFERENCE: 3759-0115P  
; CURRENT APPLICATION NUMBER: US/09/887,784  
; CURRENT FILING DATE: 2001-06-19  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequoria Victoria  
US-09-887-784-4

Query Match 99.6%; Score 1269; DB 9; Length 239;  
Best Local Similarity 99.6%; Pred. No. 2.6e-124;  
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MVSKEELFTGVVPIVLVELDGVNKGKFSVSGEGDATYGKLTGKLPVWPWT	60
Db	1	MVSKEELFTGVVPIVLVELDGVNKGKFSVSGEGDATYGKLTGKLPVWPWT	60
Qy	61	LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVYKFEQDITL	120
Db	61	LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVYKFEQDITL	120
Qy	121	VNRIELKGDIFKEDGNILGHKLEYNNSHVIMADKQKNGIKVNPKIRHNIEDGSVQLA	180
Db	121	VNRIELKGDIFKEDGNILGHKLEYNNSHVIMADKQKNGIKVNPKIRHNIEDGSVQLA	180
Qy	181	DHYQONTPTGDPVLLPDNNHYLSTOSALSQDNKRDHMLAGFVTAAGITLGMDELYK	239
Db	181	DHYQONTPTGDPVLLPDNNHYLSTOSALSQDNKRDHMLAGFVTAAGITLGMDELYK	239

```
RESULT 2
US-10-296-953-4
; Sequence 4, Application US/10296953
; Publication No. US2004007295A1
; GENERAL INFORMATION:
; APPLICANT: BJORN SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PLO095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-4

Query Match          99.6%; Score 1269; DB 12; Length 239;
Best Local Similarity 99.6%; Pred. No. 2.6e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLPKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLPKFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNNHLSQSALSKDPNEKRDHMLAGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNNHLSQSALSKDPNEKRDHMLAGFVTAAGITLGMDELYK 239

RESULT 3
US-10-270-223-6
; Sequence 5, Application US/10270223
; Publication No. US2003014363A1
; GENERAL INFORMATION:
; APPLICANT: BioImage A/S
; TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR COMPONENTS
; TITLE OF INVENTION: INTERACTING CELLS, AND TO EXTRACT QUANTITATIVE INFORMATION FROM CELLULAR COMPONENTS
; FILE REFERENCE: 3759-0126P
; CURRENT APPLICATION NUMBER: US/10/270,223
; CURRENT FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Aequorea Victoria and Human
US-10-270-223-6

Query Match          99.6%; Score 1269; DB 14; Length 363;
Best Local Similarity 99.6%; Pred. No. 4.7e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLPKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLPKFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNNHLSQSALSKDPNEKRDHMLAGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNNHLSQSALSKDPNEKRDHMLAGFVTAAGITLGMDELYK 239

RESULT 4
US-10-257-909A-30
; Sequence 30, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 893
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-30

Query Match          99.6%; Score 1269; DB 14; Length 893;
Best Local Similarity 99.6%; Pred. No. 1.7e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLPKFICTTGKLPVPWPT 60
DB 655 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLPKFICTTGKLPVPWPT 714
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 715 LVTTLISYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 774
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
DB 775 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 834
QY 181 DHYQONTPIGDGPVLLPDNNHLSQSALSKDPNEKRDHMLAGFVTAAGITLGMDELYK 239
DB 835 DHYQONTPIGDGPVLLPDNNHLSQSALSKDPNEKRDHMLAGFVTAAGITLGMDELYK 893

RESULT 5
US-10-257-909A-32
; Sequence 32, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLPKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLPKFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNNHLSQSALSKDPNEKRDHMLAGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNNHLSQSALSKDPNEKRDHMLAGFVTAAGITLGMDELYK 239

RESULT 4
US-10-257-909A-30
; Sequence 30, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 893
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-30

Query Match          99.6%; Score 1269; DB 14; Length 893;
Best Local Similarity 99.6%; Pred. No. 1.7e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLPKFICTTGKLPVPWPT 60
DB 655 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLPKFICTTGKLPVPWPT 714
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 715 LVTTLISYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 774
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
DB 775 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 834
QY 181 DHYQONTPIGDGPVLLPDNNHLSQSALSKDPNEKRDHMLAGFVTAAGITLGMDELYK 239
DB 835 DHYQONTPIGDGPVLLPDNNHLSQSALSKDPNEKRDHMLAGFVTAAGITLGMDELYK 893

RESULT 5
US-10-257-909A-32
; Sequence 32, Application US/10257909A
; Publication No. US20030187056A1
; GENERAL INFORMATION:
; APPLICANT: Bernard R. TERRY et al.
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes
; FILE REFERENCE: 3759-0125P
; CURRENT APPLICATION NUMBER: US/10/257,909A
; CURRENT FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 32
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-32

Query Match          99.6%; Score 1269; DB 14; Length 1132;
Best Local Similarity 99.6%; Pred. No. 2.4e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
DB 894 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 953

QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 954 LVTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 1013

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNPKIRHNIEDGSVQLA 180
DB 1014 VNRIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNPKIRHNIEDGSVQLA 1073

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLAGFVTAAGITLGMDELYK 239
DB 1074 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 1132

RESULT 6
US-09-887-784-2
; Sequence 2, Application US/09887784
; Patent No. US20020177189A1
; GENERAL INFORMATION:
; APPLICANT: BJOERN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea Victoria
US-09-887-784-2

Query Match          99.0%; Score 1261; DB 9; Length 239;
Best Local Similarity 99.2%; Pred. No. 1.8e-123;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60

QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNPKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNPKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLAGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 7
US-10-296-953-2
; Sequence 2, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
```

```
; APPLICANT: BJOERN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-03-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match          99.0%; Score 1261; DB 12; Length 239;
Best Local Similarity 99.2%; Pred. No. 1.8e-123;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60

QY 61 LVTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNPKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNPKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLAGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 8
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. US20020083488A1
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match          98.7%; Score 1258; DB 9; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.6e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
```

QY 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLAFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLAFVTAAGITLGMDELYK 239

RESULT 9  
US-09-999-745-4  
; Sequence 4, Application US/09999745  
; Patent No. US20020157120A1  
; GENERAL INFORMATION:  
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Baird, Geoffrey  
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS  
; FILE REFERENCE: REGEN1470-1  
; CURRENT APPLICATION NUMBER: US/09/999,745  
; CURRENT FILING DATE: 2001-10-23  
; PRIOR FILING DATE: 1999-05-21  
; NUMBER OF SEQ ID NOS: 67  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
US-09-999-745-4

Query Match 98.7%; Score 1258; DB 9; Length 239;  
Best Local Similarity 98.7%; Pred. No. 3.6e-123;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60  
QY 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLAFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLAFVTAAGITLGMDELYK 239

RESULT 10  
US-09-866-538-4  
; Sequence 4, Application US/09866538  
; Publication No. US20030032089A1  
; GENERAL INFORMATION:  
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA  
; APPLICANT: Campbell, Robert  
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS  
; FILE REFERENCE: REGEN1530-2  
; CURRENT APPLICATION NUMBER: US/09/866,538  
; CURRENT FILING DATE: 2001-05-24  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 239  
; TYPE: PRT

; ORGANISM: Aequorea victoria  
US-09-866-538-4

Query Match 98.7%; Score 1258; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 3.6e-123;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60  
QY 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLAFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLAFVTAAGITLGMDELYK 239

RESULT 11  
US-09-797-496B-2  
; Sequence 2, Application US/09797496B  
; Publication No. US20030049597A1  
; GENERAL INFORMATION:  
; APPLICANT: Simon, Sanford M.  
; APPLICANT: Chen, Yu  
; TITLE OF INVENTION: Chimeric Fluorescent Enzymes and Uses Thereof  
; FILE REFERENCE: 600-1-267  
; CURRENT APPLICATION NUMBER: US/09/797,496B  
; CURRENT FILING DATE: 2002-05-24  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Aequorea victoria green fluorescent protein modified as described  
; OTHER INFORMATION: in specification  
US-09-797-496B-2

Query Match 98.7%; Score 1258; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 3.6e-123;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTLPKICTTGKLPVPWPT 60  
QY 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLAFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLAFVTAAGITLGMDELYK 239

RESULT 12  
US-09-794-308-4  
; Sequence 4, Application US/09794308  
; Publication No. US20030170911A1  
; GENERAL INFORMATION:  
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA



APPLICANT: TSIEH, Roger  
APPLICANT: ZACHARIAS, David  
APPLICANT: BAIRD, Geoffrey  
TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS  
CURRENT APPLICATION NUMBER: US/09/794,308  
CURRENT FILING DATE: 2001-02-26  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
US-09-794-308-4

Query Match 98.7%; Score 1258; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 3.6e-123;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDGYGKLTAKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDGYGKLTAKFICTTGKLPVPWPT 60  
QY 61 LVTLLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDPFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDPFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLAGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239

RESULT 13  
US-09-865-291-4  
Sequence 4, Application US/09865291  
Publication No. US20030186229A1  
GENERAL INFORMATION:  
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA  
APPLICANT: TSIEH, Roger  
APPLICANT: TING, Alice  
APPLICANT: ZHANG, Jin  
TITLE OF INVENTION: EMISSION RATIONETRIC INDICATORS OF PHOSPHORYLATION  
FILE REFERENCE: REGEN1550  
CURRENT APPLICATION NUMBER: US/09/865,291  
CURRENT FILING DATE: 2001-05-24  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
US-09-865-291-4

Query Match 98.7%; Score 1258; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 3.6e-123;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDGYGKLTAKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDGYGKLTAKFICTTGKLPVPWPT 60  
QY 61 LVTLLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDPFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDPFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLAGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239

RESULT 14  
US-10-457-982-3  
Sequence 3, Application US/10457982  
Publication No. US2003021265A1  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger Y.  
APPLICANT: Miyawaki, Atsushi  
APPLICANT: Llopis, Juan  
APPLICANT: Wachter, Rebekka M.  
APPLICANT: Remington, S. James  
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
FILE REFERENCE: 07257/071001  
CURRENT APPLICATION NUMBER: US/10/457,982  
CURRENT FILING DATE: 2003-06-09  
PRIOR APPLICATION NUMBER: US/09/602,641  
PRIOR FILING DATE: 2000-06-22  
PRIOR APPLICATION NUMBER: 09/172,063  
PRIOR FILING DATE: 1998-10-13  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (0)...(0)  
OTHER INFORMATION: EGFP  
US-10-457-982-3

Query Match 98.7%; Score 1258; DB 12; Length 239;  
Best Local Similarity 98.7%; Pred. No. 3.6e-123;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDGYGKLTAKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDGYGKLTAKFICTTGKLPVPWPT 60  
QY 61 LVTLLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDPFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDPFKEDGNILGHKLEYNHNYSNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLAGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239

RESULT 15  
US-10-121-258-13  
Sequence 13, Application US/10121258  
Publication No. US20030059835A1  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger  
APPLICANT: Campbell, Robert  
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT  
TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME  
FILE REFERENCE: UC083.1CP2CP1  
CURRENT APPLICATION NUMBER: US/10/121,258  
CURRENT FILING DATE: 2002-04-10  
PRIOR APPLICATION NUMBER: 09/794,308  
PRIOR FILING DATE: 2001-02-26  
PRIOR APPLICATION NUMBER: 09/866,538  
PRIOR FILING DATE: 2001-05-24

```
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Enhanced Green Fluorescent Protein (EGFP)
US-10-121-258-13

Query Match      98.7%; Score 1258; DB 14; Length 239;
Best Local Similarity 98.7%; Pred. NO. 3.6e-123;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 MVSKEBELFTGVVPIILVELDGVNGHKFSVSGEGDGYGKLTLEFICTTGKLPVPWPT 60
Db      1 MVSKEBELFTGVVPIILVELDGVNGHKFSVSGEGDGYGKLTLEFICTTGKLPVPWPT 60

QY     61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db     61 LVTTLYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY     121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db     121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180

QY     181 DHYQONTPIGCGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGFVTAAGITLGMDELYK 239
Db     181 DHYQONTPIGCGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:09:30
Job time : 35.7778 secs
```

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	1236	97.0	238	1	JQ5114	green-fluorescent
2	106	8.3	785	2	H73228	hypothetical prote
3	91.5	7.2	861	2	H64102	leucine-tRNA ligas
4	90	7.1	632	2	T06586	DNA-binding protei
5	89.5	7.0	887	2	E82590	leucyl-tRNA synthet
6	89	7.0	578	1	I40794	dihydrodipolamide d
7	88.5	6.9	655	2	D83917	DNA topoisomerase
8	87.5	6.9	370	2	E70390	iron-sulfur cofact
9	87.5	6.9	860	2	AC0582	leucyl-tRNA synthet
10	87.5	6.9	222	1	A36028	DNA-directed DNA p
11	87.5	6.9	257	2	D71614	hypothetical prote
12	86	6.8	357	2	G81355	trRNA (uracil-5)-am
13	86	6.8	461	2	T06936	photosystem II chi
14	85.5	6.7	788	1	JCVLHH	DNA-directed DNA p
15	85.5	6.7	797	2	JC4078	protective surface
16	85.5	6.7	808	2	F64102	protective surface
17	85.5	6.7	889	2	JC5576	inter-alpha-trypsi
18	85	6.7	281	2	AD2052	hypothetical prote
19	84.5	6.6	310	2	S68225	synergonemotropi
20	84.5	6.6	425	2	C97354	hypothetical prote
21	84.5	6.6	613	2	A99552	oligodeopeptidase
22	84.5	6.6	941	2	S29043	cellulase (EC 3.2.
23	84	6.6	353	2	E84941	imidazoleglycerol-
24	84	6.6	874	2	JC4930	S-layer protein pr
25	83.5	6.6	263	2	S53488	water-stress-induc
26	83.5	6.6	534	1	NICUMA	nitrogenase (EC 1.
27	83.5	6.6	836	1	JDVLDD	DNA-directed DNA p
28	83.5	6.6	1134	2	A60234	IgA Fc receptor pr
29	83.5	6.6	1164	1	FC50AG	IgA Fc receptor pr

RESULT 3  
H64102  
leucine-tRNA ligase (EC 6.1.1.4) - Haemophilus influenzae (strain Rd KW20)  
Alternate names: leucyl-tRNA synthetase  
C.Species: Haemophilus influenzae  
C.Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 03-Jun-2002  
C.Accession: H64102  
R.Fleischmann, R.D. Adams, M.D. White, O. Clayton, R.A.; Kirkness, E.F.;  
Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.;  
Futerman, J.L.; Roberts, J.; White, O.; Adams, M.D.; White, O.; Clayton, R.A.;

Science 269, 496-512, 1995  
A; Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.A.; Title: Whole-genome random sequencing and assembly of *Haemophilus influenzae* type 26400. PMID: 7542800

A;Accession: H64102  
A;Status: nucleic acid sequence not shown; translation not shown

A;Residues: 1-861 <PIGR>  
A;Cross-references: GB:U32774; GB:U32774; GB:L42023; NID:g1573942; PIDN:AAC2581.1; PI

A; Gene: leuS  
C; Superfamily: leucine-tRNA ligase  
C; Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match	7.2%;	Score 91.5;	DB 2;	Length 861;
Best Local Similarity	24.1%;	Pred. No. 14;		

50 TTGKLPVPWPTLVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDD-----

Db 314 TGDKLPI-WVANFVLMHYGTGAVMAVPAH-DQDF-----EFAQKYSLPKQVIAPLA

Db 365 DEEIDLTKQAFVEHGKLVNSDEFDGKNF--DGAFN-----IADKLEKLGV

```
QY      163 ---KVNFKIRH-----NIEDGSVQLADHYQONTPTAGDGVVLDFDNRIID-  
          :|::|||:         :|::|||:         :|::|||:         :|::|||:  
          -----DMEDLPITIPEDVVMD
```

QY 203 STQSALSKDPN 213

Db 462 GVKSPINADPN 472

RESULT 4  
T08586  
ova binding protein Dn3 - garden pea

C:species: Fisum sativum (garbanzo pea),  
C:date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 08-Oct-1999  
C:accession: T06586

A-description: Identification of a novel family of DNA-binding proteins with  
submitted to the EMBL Data Library, June 1996

A;Accession: T06586  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Reference number: E13771

A;Residues: 1-632 <SAT>  
A;Cross-references: EMBL:X98740; NID:e995229; PIDN:CAA67292.1; PID:e275185

Query Match	7.1%	Score 90	DB 2	Length 632

Matches 49; Conservative 26; Mismatches 19; Indels 36; Gaps 36

```
db      IVDGRDVGSKVDVINKESNEATIPENK-----PTEPKLDVEQLAATTM-----
363
```

QY 76 PDHMKQHDFFKSAMPEQIV-----QENRIIRADDDGRNRIAGC-R-----

Db 409 PSSAKNVLTOLIVFTPLRSVARTSGREGSEELKDSGNSLERDTKKLELEQGKNS-E 467  
QY 126 LKIDFKEDGNILGHKLE-----YYNSHNVYIMADKQK-----N 160  
Db 468 LKGIETDNTSLDEKFNALGNKILKEISNPRHDVESANSHTHNQVTVSHQKALETN 527  
QY 161 GIKVNFKIRHNIEDG-----SVQLADHYQ 194  
Db 528 QSQVEDVAKNIQDDSKPSESLHAKDYR 557

RESULT 5  
E82590  
leucyl-tRNA synthetase XF2176 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
C:Accession: E82590  
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717; PMID:10910347  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: E82590  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-887 <SIM>  
A:Cross-references: GB:AE004031; GB:AE003849; NID:g9107309; PIDN:AAF84975.1; GSPDB:GN001  
A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H  
as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laigh  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.B.; Marques, M.V.; Martins, E  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF2176  
C:Superfamily: leucine-tRNA ligase

Query Match 7.0%; Score 89.5; DB 2; Length 887;  
Best Local Similarity 22.2%; Pred. No. 21;  
Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;  
QY 50 TTGKLPVWPPTLVTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNY-- 107  
Db 329 TNEQLPV-WVANFVLMAYGTGAVMAYVGHQDQDEF--ANKYGLPIRQVIALKEPKNQDB 385  
QY 108 -----KTRAEVKFPGDTLVNRIELKLGIDFKEDGNILGHKLEYNVSNVYI 153  
Db 386 STWEPDWRDWDADKTR---EFE---LINSAEFDGLDYQDAEVLAEPE----- 429  
QY 154 MADKQKNG-IKVNFKIRHNIEDGSVLADHYQQNTPI-----GGPVLPLPDN 199  
Db 430 ---RQGRQRVNYRLR-----DWGVSQRQYWGCPVPVICTGAVPVPEDQLPVILPEN 482  
QY 200 -HYLSTQSALSCKDPNEKR 216  
Db 483 VAFSGTGSPKTPDPEWRK 500

RESULT 6  
I40794  
dihydrolipoamide dehydrogenase (EC 1.8.1.4) [validated] - Clostridium magnum  
N:Alternative names: 2-oxoglutarate dehydrogenase complex chain E3; acetoin dehydrogenase  
hydrogenase complex chain E3; S-complex 50K chain  
C:Species: Clostridium magnum

Db 409 PSSAKNVLTOLIVFTPLRSVARTSGREGSEELKDSGNSLERDTKKLELEQGKNS-E 467  
QY 126 LKIDFKEDGNILGHKLE-----YYNSHNVYIMADKQK-----N 160  
Db 468 LKGIETDNTSLDEKFNALGNKILKEISNPRHDVESANSHTHNQVTVSHQKALETN 527  
QY 161 GIKVNFKIRHNIEDG-----SVQLADHYQ 194  
Db 528 QSQVEDVAKNIQDDSKPSESLHAKDYR 557

RESULT 5  
E82590  
leucyl-tRNA synthetase XF2176 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
C:Accession: E82590  
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717; PMID:10910347  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: E82590  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-887 <SIM>  
A:Cross-references: GB:AE004031; GB:AE003849; NID:g9107309; PIDN:AAF84975.1; GSPDB:GN001  
A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H  
as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laigh  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.B.; Marques, M.V.; Martins, E  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF2176  
C:Superfamily: leucine-tRNA ligase

Query Match 7.0%; Score 89.5; DB 2; Length 887;  
Best Local Similarity 22.2%; Pred. No. 21;  
Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;  
QY 50 TTGKLPVWPPTLVTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNY-- 107  
Db 329 TNEQLPV-WVANFVLMAYGTGAVMAYVGHQDQDEF--ANKYGLPIRQVIALKEPKNQDB 385  
QY 108 -----KTRAEVKFPGDTLVNRIELKLGIDFKEDGNILGHKLEYNVSNVYI 153  
Db 386 STWEPDWRDWDADKTR---EFE---LINSAEFDGLDYQDAEVLAEPE----- 429  
QY 154 MADKQKNG-IKVNFKIRHNIEDGSVLADHYQQNTPI-----GGPVLPLPDN 199  
Db 430 ---RQGRQRVNYRLR-----DWGVSQRQYWGCPVPVICTGAVPVPEDQLPVILPEN 482  
QY 200 -HYLSTQSALSCKDPNEKR 216  
Db 483 VAFSGTGSPKTPDPEWRK 500

RESULT 6  
I40794  
dihydrolipoamide dehydrogenase (EC 1.8.1.4) [validated] - Clostridium magnum  
N:Alternative names: 2-oxoglutarate dehydrogenase complex chain E3; acetoin dehydrogenase  
hydrogenase complex chain E3; S-complex 50K chain  
C:Species: Clostridium magnum

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
C:Accession: I40794  
R:Kruger, N.; Opermann, F.B.; Lorenzl, H.; Steinbuchel, A.  
J. Bacteriol. 176, 3614-3630, 1994  
A:Title: Biochemical and molecular characterization of the Clostridium magnum acetoin de  
A:Reference number: I40789; MUID:94266715; PMID:8206840  
A:Accession: I40794  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-578 <RU>  
A:Cross-references: GB:IJ1844; NID:g472324; PIDN:AAA21748.1; PID:g472330  
C:Function:  
A:Description: catalyzes the oxidation of dihydrolipoamide to lipoamide using NAD  
A:Pathway: acetoin dehydrogenase enzyme system  
C:Superfamily: Alkaligenes dihydrolipoamide dehydrogenase; dihydrolipoamide dehydrogenase  
C:Keywords: FAD; flavoprotein; lipoamide; NAD; oxidoreductase; redox-active disulfide  
F15-77/Domain: lipoyl/biotin-binding homology <LBP>  
F117-145/Region: beta-alpha-beta FAD nucleotide-binding fold  
F119-561/Domain: dihydrolipoamide dehydrogenase homology <DL>  
F1287-315/Region: beta-alpha-beta NAD nucleotide-binding fold  
F153-158/Disulfide bonds: redox-active #status predicted

Query Match 7.0%; Score 89; DB 1; Length 578;  
Best Local Similarity 22.9%; Pred. No. 13;  
Matches 57; Conservative 43; Mismatches 91; Indels 58; Gaps 13;  
QY 10 TGVVPTLVLDGVNGHKFSVSGEGDATYKGLTKL-----FICTGKLPVWPPTLVT 63  
Db 255 TGSMPFIPBIE-----GNKLS-----GVIST-GALSLESNPESIALGGVIGVEFASIFN 305  
QY 64 TLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFPGDTLVNR 123  
Db 306 SLGCKSVIIEMLPHILPMDREISEI-----AKAKLIRDGININN 346  
QY 124 ILKGIIDFKEDG---NILGHKLEYNVSNVIMADKQK---GIKNFKIRHNIEDGSVQ 178  
Db 347 CKVTRIEQGEDGLKVSFIGDKGEESIDVEKVLIAVGRSNIEGLDVE-KIGVKTEGSGII 405  
QY 179 LADHYQQNT---PIGD--GPVLLPDNHYLSTQSALSCKD---PNEKRDMHVLGAFV--- 225  
Db 406 VNDKMETNVEGIYAIGDCTGKIMLA--HVASQGVVAENINGQNKMDYKTVPACVYTK 463  
QY 226 ---TAAGIT 231  
Db 464 PELASVGLT 472

RESULT 7  
D83917  
DNA topoisomerase IV subunit B BH2140 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C:Accession: D83917  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, F.; Hirai  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MUID:20512592; PMID:11058132  
A:Accession: D83917  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-655 <STO>  
A:Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BA05859.1; GSPDB:GN001  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH2140  
C:Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B

Query Match 6.9%; Score 88.5; DB 2; Length 655;  
Best Local Similarity 21.9%; Pred. No. 17;  
Matches 53; Conservative 37; Mismatches 95; Indels 57; Gaps 9;  
QY 22 DVNGHK---FSVSGEGEGDAT----YGLTKLTKFI-----CTTGKLPVWP 59

```
Db      63 NVTIHKDQSVSRDEGRGMPGTMHKLKPTPEVILTVLHAGGKFGQGGVATSGLHGVA 122
QY      60 TLVTTLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQER-----TIFPKDQG----- 105
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      123 SVYNALSEWLIVIKRDGWYEQRFENGKGPSTLEKKGKTROTGTTHFKPDPVTFSTT 182
QY      106 --NFKTRAEVKFGDGLVNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADK----- 157
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      183 NFNVETLSELRRAAFLKGLKLELDVLRDRTKEVPH-YEDGKAQVEYLNEDKETLHPV 241
QY      158 -----QKNGIKVNFKIRHNIEDSGVOLADHYQOQTPIGDGFVLLPNHNYLSTQSALSKDP 212
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      242 VFENGESNGLEIEFAQFN--DGYTENVLVSFVNVNRTKDG-----GTHELGAKTANTRAV 294
QY      213 NE 214
       ||
Db      295 NE 296

RESULT 8
E70390
Iron-sulfur cofactor synthesis protein nifs - Aquifex aeolicus
N;Contains: L-cysteine sulfurtransferase (EC 2.8.1.-)
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 07-Dec-1999
C;Accession: E70390
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: E70390
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-370 <AQF>
A;Cross-references: GB:AE000720; NID:g2983529; PIDN:AAC07111.1; PID:g2983536; GB:AE00065
A;Experimental source: strain VFS
C;Genetics:
C;Superfamily: nitrogen fixation protein nifs
C;Keywords: phosphoprotein; pyridoxal phosphate; sulfurtransferase
F;195/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted
F;318/Active site: Cys (cysteine persulfide intermediate) #status predicted

Query Match      6.9%; Score 87.5; DB 2; Length 370;
Best Local Similarity 25.4%; Pred. No. 9.9;
Matches 49; Conservative 30; Mismatches 91; Indels 23; Gaps 8;

QY      4 KGBELFTGVV-----PILVELD----GDVNGHKE-SVSGEG-----EGDATYGLKTLKFICT 50
Db      164 KGVLPLTDAVQATGKPIELKNISYATFSGKHKPHAIXGSLFLISDEANVEPLIVGGQE 223
QY      51 TGKLP-----VPMPTLTVTLLSYGVQCFSRYPDHMKQ-HDFFKSAMPEGYVQERTIFFKOD 104
Db      224 NGRKSGTENNVGILSLAKALEIIVNFSRYOEQLKKLRDLFNLLLEA-LPDAQIVGKDA 282
QY      105 GNVKTRAEV---KPEGDTLVNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNG 161
Db      283 ERSPSISVVIMPKFFGAIVNKLSEKGIYCTSGSACLSGSEYEPNKMVKMGFSQEXALRM 342
QY      162 IKVNFKIRHNIED 174
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      343 VRFSFGLLNKEE 355

RESULT 9
AC0582
leucyl-tRNA synthetase [imported] - Salmonella enterica subsp. enterica serovar Typhi (e
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AC0582
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
```

```
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AC0582
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-860 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD05125.1; PID:g16501899; GSPDB:GN00176
C;Genetics:
A;Gene: STY0699
C;Superfamily: leucine-tRNA ligase

Query Match      6.9%; Score 87.5; DB 2; Length 860;
Best Local Similarity 23.3%; Pred. No. 30;
Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;

QY      50 TTCKLPVPMPTLTVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKT 109
Db      314 TGEETIPV-WAANFVLMYEYGTGAVMVPGH-DORD-YEFASKYGLTIKPVILAADGSEPD 370
QY      110 RAEVKEGDTLVNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIR 169
Db      371 SEQALTEKGVLFNSGFEFGDLAEAFNAIADKL-----AEKGVGERKVNRLR 418
QY      170 H-----NTEDSVQLADHYQOQTPIGDGFVLLPNHNYL-STQSALSKDP 212
Db      419 DMGVSQRQYWGAPIMVTLIEDGTV-----LPTPDQLPEVLDPEDVMDGITSPIKADP 471

RESULT 10
A36028
DNA-directed DNA polymerase (BC 2.7.7.7) II catalytic chain - yeast (Saccharomyces cerevisi
N;Alternate names: DNA-directed DNA polymerase II chain A; protein N0825; protein YNL262v
C;Species: Saccharomyces cerevisiae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: A36028; B36028; S60919; S63235; S65121
R;Morrison, A.; Arai, H.; Clark, A.B.; Hamatake, R.K.; Sugino, A.
Cell 62, 1143-1151, 1990
A;Title: A third essential DNA polymerase in Saccharomyces cerevisiae.
A;Reference number: A36028; MUID:90381771; PMID:2169349
A;Accession: A36028
A;Molecule type: DNA
A;Residues: 1-2222 <MOR>
A;Cross-references: GB:M60416; GB:M36724; NID:gl71408; PIDN:AAA88711.1; PID:gl71409
A;Accession: B36028
A;Molecule type: protein
A;Residues: 1214-1216, 'X', 1218-1221 <MO2>
R;Sen-Gupta, M.; Lyck, R.; Fleig, U.; Niedenthal, R.N.; Hegemann, J.M.
submitted to the EMBL Data Library, October 1995
A;Description: The sequence of a 24152 bp segment from the left arm of chromosome XIV fr
A;Reference number: S60909
A;Accession: S60919
A;Molecule type: DNA
A;Residues: 1-2221 <SEN>
A;Cross-references: EMBL:X92494; NID:g1045236; PIDN:CAA63235.1; PID:g1045247
R;Sen-Gupta, M.; Lyck, R.; Niedenthal, R.K.; Fleig, U.N.; Hegemann, J.H.
submitted to the Protein Sequence Database, April 1996
A;Reference number: S63235
A;Accession: S63235
A;Molecule type: DNA
A;Residues: 1-2222 <SEW>
A;Cross-references: EMBL:Z71538; NID:gl302316; PIDN:CAA96169.1; PID:gl302317; GSPDB:GN000
A;Experimental source: strain S288C
R;Sen-Gupta, M.; Lyck, R.; Fleig, U.; Niedenthal, R.K.; Hegemann, J.H.
Yeast 12, 505-514, 1996
A;Title: The sequence of a 24 152 bp segment from the left arm of chromosome XIV from Sac
A;Reference number: S65111; MUID:96310631; PMID:8740425
A;Accession: S65121
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
```

A;Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyp

A;Reference number: AB1250; MUID:20150912; PMID:10688204

A;Accession: G81355

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-357 <PAR>

A;Cross-references: GB:AL139076; GB:AL111168; NID:G6968128; PIDN:CAB73096.1; PID:G696827;

A;Experimental source: serotype O2, strain NCTC 11168

C;Genetics:

A;Gene: trmA; Cj0831c

C;Keywords: methyltransferase; S-adenosylmethionine

Query Match 6.8%; Score 86; DB 2; Length 357;  
Best Local Similarity 24.8%; Pred. No. 13; Mismatches 39; Indels 34; Gaps 5;  
Matches 30; Conservative 18;

QY 80 KOHDFFSAMPEGVQVQERTTFFKDDGNYKTRAEVKF--EGDTLV-----NRIELKG 128  
DB 14 EKHSFIKKYKPEFTTKDFKLPASKDKHYTRAELSFYHENDTLFYAMPDPKSKKVIIEV 73  
QY 129 IDFKED-----GNILGHKLEYNYNSHNVYIMADKQNGIKVNFKIRHNIE 173  
DB 74 LDFADEKICAFMPRLLEVLQRDNKLEKEK-----FGVEFLTQKE--LSITLLYHKNIE 125  
QY 174 D 174  
DB 126 D 126

RESULT 13

T06936

photosystem II chlorophyll a-binding protein psbC - Cyanophora paradoxa cyanelle  
N;Alternate names: protein CP43  
C;Species: cyanelle Cyanophora paradoxa  
C;Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 26-Aug-1999  
C;Accession: T06936  
R;Stirrewalt, V.L.; Michalowski, C.B.; Luffelhardt, W.; Bohnert, H.J.; Bryant, D.A.  
submitted to the EMBL Data Library, July 1995  
A;Description: Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*.  
A;Reference number: Z15840  
A;Accession: T06936  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-461 <STR>  
A;Cross-references: EMBL:U30821; NID:g1016083; PIDN:AAA81279.1; PID:g1016192  
A;Experimental source: strain Pringsheim LB555  
C;Genetics:  
A;Gene: psbC  
A;Genome: cyanelle  
C;Superfamily: photosystem II chlorophyll a-binding protein psbC  
C;Keywords: chlorophyll; cyanelle; membrane-associated complex; photosynthesis; photosys

Query Match 6.8%; Score 86; DB 2; Length 461;  
Best Local Similarity 21.0%; Pred. No. 18; Mismatches 89; Indels 90; Gaps 13;  
Matches 57; Conservative 35;

QY 25 GHKFSVS6GEGDATYKGLTLKFICTTGKL-----PVPWP--TLVTT----LSYGV--- 69  
DB 208 GDGHWISVDNMEDIIGHIWLAFICIIGVWHLLTKPFSWARRALVWSEAYLSYLAAL 267  
QY 70 -----QCFSRYPDHMKQHDFFKSAMPE-GVQERTTFFKDD-----GN 106  
DB 268 ALMGFIANCFVWFNTATYPSEFFGTGPEASQAQAFTELVRDQLGANVGSAGQPTGLGK 327  
QY 107 YKTRF---EVKPEGDTLVNRLELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQNGIK 163  
DB 328 YLMKPSGEIIFGGETM-----RFDWTRAPWLEPLRGANGLD 364  
QY 164 VNFKIRNIEDGSVOLADHYQOQNTPIGD---GPVLLPDN--HYLSTQSALSKDPNEKRD 217  
DB 365 LT-KIKYDIQWQERRAAEYMTWAPLGSLSNVGGVATINSVNYVSPRSWLS-----TS 417  
QY 218 HMVLAGEFV-----TAAGITLGM 235

```

487 FFENYDMSKSDTSNYKRTTGYGNSVTL-GPPVNNNSYYVGLGHTYKNISNFALYN---542
Db

149 HNVYIMADKQK-NGIKVNFKIRINIEDGSQLADHYQQ-----NTPDGDGPVLL196
Qy

543 RNLVYQSMKEFGNGIKTN-----DFDFSGWNVNSLNRGYFPTGKVKASLG-GRVTI593
Db

197 P--DNHYLSTQSLSKDPEKRDHMLVLAGVTTAAGITLG233
Qy

594 PGSDNKYTKLSADVQGFYPLDRDHLWVWSAKASAGYANG632
Db

```



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:42:24 ; Search time 6.44444 Seconds  
(without alignments)  
1931.085 Million cell updates/sec

Title: US-09-887-784-222A

Perfect score: 1274

Sequence: 1 MVSKEELFTGVVPIVELD.....VLAQFVTAAGITLGMDELK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1242	97.5	238	1	GFP_AEQVI
2	91.5	7.2	861	1	SYL_HAEN
3	89.5	7.0	879	1	SYL_XYLPA
4	87.5	6.9	860	1	SYL_SALTI
5	87.5	6.9	860	1	SYL_SALTY
6	87.5	6.9	2222	1	DPOE_YEAST
7	87	6.8	689	1	AC2L_HUMAN
8	86.5	6.8	533	1	CP51_CANGA
9	86.5	6.8	879	1	SYL_XYLFT
10	86.5	6.8	1603	1	VITA_CARL
11	86	6.8	357	1	TRMA_CAMJE
12	86	6.8	461	1	PSBC_CRAPA
13	86	6.8	504	1	YC03_KLEPN
14	85.5	6.7	501	1	AMPA_WIGBR
15	85.5	6.7	788	1	DPOI_HPBHE
16	85.5	6.7	795	1	D152_HAEN
17	85.5	6.7	797	1	D151_HAEN
18	85.5	6.7	886	1	ITH_MESAU
19	84.5	6.6	589	1	SYD_HAEDU
20	84.5	6.6	613	1	PEPF_MYCPU
21	84.5	6.6	793	1	D153_HAEN
22	84.5	6.6	859	1	SYL_SHRON
23	84.5	6.6	941	1	GUN_BACS6
24	84	6.6	353	1	HIS7_BUCAI
25	84	6.6	366	1	SET7_HUMAN
26	84	6.6	874	1	SLAP_BACLI
27	83.5	6.6	533	1	NIFD_CLOPA
28	83.5	6.6	538	1	GRBE_RAT
29	83.5	6.6	1184	1	BAG_STRAG
30	82	6.4	336	1	YD4B_METJA
31	82	6.4	461	1	PSBC_CHLEU
32	82	6.4	682	1	PRC_ECOLI
33	82	6.4	737	1	OPTI_DROME

RESULT 1									
GFP_AEQVI									
ID	GFP_AEQVI	STANDARD;		PRT;	238 AA.				
AC	P42212; Q17104; Q27903;								
DT	01-NOV-1995 (Rel. 32, Created)								
DT	01-NOV-1995 (Rel. 32, Last sequence update)								
DT	10-OCT-2003 (Rel. 42, Last annotation update)								
DE	Green fluorescent protein.								
GN	GFP.								
OS	Aequorea victoria (Jellyfish).								
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;								
OC	Aequoreidae; Aequorea.								
OX	NCBI_TaxID=6100;								
RN	[1]								
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.								
RX	MEDLINE=92175527; PubMed=1347277;								
RA	Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,								
RA	Cormier M.J.;								
RT	"Primary structure of the Aequorea victoria green-fluorescent								
RT	protein.";								
RL	Gene 111:229-233 (1992).								
RN	[2]								
RP	SEQUENCE FROM N.A.								
RX	MEDLINE=94185810; PubMed=8137953;								
RA	Inouye S., Tsuji F.I.;								
RT	"Aequorea green fluorescent protein. Expression of the gene and								
RT	fluorescence characteristics of the recombinant protein.";								
RL	FEBS Lett. 341:277-280 (1994).								
RN	[3]								
RP	SEQUENCE FROM N.A.								
RX	MEDLINE=97299832; PubMed=9154981;								
RA	Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;								
RT	"Enhanced expression in tobacco of the gene encoding green fluorescent								
RT	protein by modification of its codon usage.";								
RL	Plant Mol. Biol. 33:989-999 (1997).								
RN	[4]								
RP	CHROMOPHORE.								
RX	MEDLINE=93192221; PubMed=8448132;								
RA	Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;								
RT	"Chemical structure of the hexapeptide chromophore of the Aequorea								
RT	green-fluorescent protein.";								
RL	Biochemistry 32:1212-1218 (1993).								
RN	[5]								
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).								
RX	MEDLINE=96355665; PubMed=8703075;								
RA	Ormos M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,								
RA	Remington S.J.;								
RT	"Crystal structure of the Aequorea victoria green fluorescent								
RT	protein.";								
RL	Science 273:1392-1395 (1996).								
RN	[6]								
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).								
RX	MEDLINE=98294543; PubMed=9631087;								
RA	Yang F., Moss L.G., Phillips G.N. Jr.;								
RT	"The molecular structure of green fluorescent protein.";								
RL	Nat. Biotechnol. 14:1246-1251 (1996).								

P28840 rattus norv  
O25443 helicobacte  
Q99ub1 mus musculu  
Q42918 schizosacch  
O00116 homo sapien  
Q8xbn8 escherichia  
Q8fjy9 escherichia  
P07813 escherichia  
P39321 escherichia  
P24506 discopyge o  
P33563 bacillus eu  
Q92405 aspergillus



Qy	104	DGNKYTRAEVKEPGDTLVNRIELKGIIDFKEDGNIHLGHLEYNVSHNVYIMADK-QKNGI	167
Db	365	DBEIDLTQKAQFVEHGHKLVSDFDGKNF--DGAENG-----IADKLEKLGV	408
Qy	163	---KNVFKRHR-----NIEDSGVLADHYQQNTPTIGDGPVLIPDNHYL-	202
Db	409	GKRVNYRLRDWGSQRWYGAPIMPLTILGNDGVPA-----PMEDTLPIIPDDVMD	461
Qy	203	STQSALS KDPN 213	
Db	462	GVKSPINADPN 472	

RESULT 3  
SYL\_XYLFA

ID	SYL_XYLFA	STANDARD;	PRT; 879 AA.
AC	Q9PBG8;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DT	LEUCYL-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeURS).		
DE	GENE OR XF2176.		
OS	Xylella fastidiosa.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;		
OC	Xanthomonadaceae; Xylella.		
NCBI	TaxID=2371;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=9a5C;		
RX	MEDLINE=20365717; PubMed=10910347;		
RA	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,		
RA	Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,		
RA	Bartos M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,		
RA	Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,		
RA	Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,		
RA	Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,		
RA	Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,		
RA	Praga J.S., Francis S.C., Franco M.C., Frohme M., Furian L.R.,		
RA	Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,		
RA	Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,		
RA	Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,		
RA	Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,		
RA	Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,		
RA	Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,		
RA	Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,		
RA	Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,		
RA	Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,		
RA	de Oliveira M.C., de Oliveira R.C., Palmeri D.A., Paris A.,		
RA	Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,		
RA	Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,		
RA	de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,		
RA	da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,		
RA	da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,		
RA	de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchiko M.H.,		
RA	Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,		
RA	Zago M.A., Zatz M., Meidanis J., Setubal J.C.;		
RT	"The genome sequence of the plant pathogen Xylella fastidiosa.";		
CC	Nature 406.151-159(2000)		
CC	-!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +		
CC	diphosphate + L-leucyl-tRNA(Leu).		
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic.		
CC	-!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the ENBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
DR	EMBL; AF0004031; AAF84975.1; ALT_INIT.		
DR	HMAP; NB 00049; -; 1.		



```
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE008725; AAL19599.1; -.
CC STyGene; SG????; leus.
CC HAMAP; MF 00049; -.
CC InterPro; IPR002302; Leu-tRNA-synt1a.
CC InterPro; IPR002300; tRNA-synt_1a.
CC InterPro; IPR001412; tRNA-synt_1.
CC InterPro; IPR009008; ValRS_1Iers_edit.
CC Pfam; PF001133; tRNA-synt_1; 1.
CC PRINTS; PR00985; tRNA-syntHLEU.
CC TIGRfam; TIGR00396; leus_bact; 1.
CC PROSITE; PS00178; AA-trNA_LIGASE_I; 1.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC SITE 42 52 "HIGH" REGION.
CC SITE 619 623 "KMSKS" REGION.
CC BINDING 622 622 ATP (BY SIMILARITY).
CC SEQUENCE 860 AA; 96985 MW; D5003584DFECCAB6 CRC64;
CC -----
Query Match 6.9%; Score 87.5; DB 1; Length 860;
Best Local Similarity 23.3%; Pred. No. 15;
Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;
QY 50 TTGKLPVMPPTLVTTLSYGVCFSRYPDHMKQDFFKSAPEGYVQERTIFFKDDGNYKT 109
DB 314 TGEIIPV-WAANFVLMYEGTGAVMAVFGH-DQRD-YEFASKYGLTIKPVILAADGSEPD 370
QY 110 RAEVKEGDTLVNRIELKIDFKEDGNILGHKLEYNHNVIMADKQNGIKVNFKR 169
DB 371 SEQALTEKGVLFNSGFBDFGLAFEAFAATADKL-----AEKGVGERKVNRYLR 418
QY 170 H-----NTEGSGVQLADHYQNTPIGDGVLPLPDNHYL-STOSALSKDP 212
DB 419 DWGVSQRQYWGAPIPWVLEDTGV-----LFTPEDQLPVILPDVMDGITSPIKADP 471
RESULT 6
ID_DPOE YEAST STANDARD; PRT; 2222 AA.
AC P21951;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA polymerase epsilon, catalytic subunit A (EC 2.7.7.7) (DNA
DE polymerase II subunit A)
GN POL2 OR DUN2 OR YNL262W OR N0825.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1214-1221.
RX MEDLINE=90381771; PubMed=2169349;
RA Morrison A., Araki H., Clark A.B., Hamatake R.K., Sugino A.;
RT "A third essential DNA polymerase in S. cerevisiae.";
RL Cell 62:1143-1151(1990).
RN [2]
RP SEQUENCE OF 1-2221 FROM N.A.
RC STRAIN=S288c / FY1679;
RX MEDLINE=96310631; PubMed=8740425;
RA Sen-Gupta M., Lyck R., Fleig U., Niedenthal R.K., Hegemann J.H.;
```

```
RT "The sequence of a 24,152 bp segment from the left arm of chromosome
RT XIV from Saccharomyces cerevisiae between the BNI1 and the POL2
RT genes."
RL Yeast 12:505-514(1996).
RN [3]
RP TEMPERATURE SENSITIVE MUTANTS.
RX MEDLINE=92164663; PubMed=1537345;
RA Araki H., Ropp P.A., Johnson A.L., Johnston L.H., Morrison A.,
RA Sugino A.;
RT "DNA polymerase II, the probable homolog of mammalian DNA polymerase
RT epsilon, replicates chromosomal DNA in the yeast Saccharomyces
RT cerevisiae."
RL EMBO J. 11:733-740(1992).
CC -1- FUNCTION: DNA POLYMERASE II PARTICIPATES IN CHROMOSOMAL DNA
CC REPLICATION.
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA] (N).
CC -1- SUBUNIT: CONSISTS OF FIVE SUBUNITS (200 kDa, 80 kDa, 34 kDa, 30
CC kDa, AND 29 kDa).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: THE DNA POLYMERASE ACTIVITY DOMAIN RESIDES IN THE
CC N-TERMINAL HALF OF THE PROTEIN, WHILE THE C-TERMINUS IS NECESSARY
CC FOR COMPLEXING SUBUNITS B AND C.
CC -1- MISCELLANEOUS: In eukaryotes there are five DNA polymerases:
CC alpha, beta, gamma, delta, and epsilon which are responsible for
CC different reactions of DNA synthesis.
CC -1- SIMILARITY: Belongs to the DNA polymerase type-B family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; MG04116; AAA88711.1; -.
CC EMBL; X92494; CAA63235.1; -.
CC EMBL; Z71538; CAA96169.1; -.
CC PIR; A36028; A36028.
CC GERMOnline; 143268; -.
CC SGD; S0005206; POL2.
CC GO; GO:0000731; P:DNA repair synthesis; IMP.
CC InterPro; IPR006172; DNA pol B.
CC InterPro; IPR006134; DNA pol B dom.
CC InterPro; IPR006133; DNA pol B_exo.
CC Pfam; PF001136; DNA pol B; 1.
CC Pfam; PF03104; DNA pol B_exo; 1.
CC SMART; SM00486; POLBG; 1.
CC PROSITE; PS00116; DNA POLYMERASE B; FALSE NEG.
CC Transferrase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; zinc-finger; Nuclear protein.
FT ZN-FING 2108 2181 POTENTIAL.
FT VARIANT 644 644 M -> I (IN POL2-9 TS MUTANT).
FT VARIANT 710 710 P -> S (IN POL2-18 TS MUTANT).
SQ SEQUENCE 2222 AA; 255669 MW; CBCDDE2AB147D65B CRC64;
Query Match 6.9%; Score 87.5; DB 1; Length 2222;
Best Local Similarity 28.2%; Pred. No. 46;
Matches 37; Conservative 14; Mismatches 49; Indels 31; Gaps 7;
QY 54 LPVPWP-TLVTTLSYGVCFSRYPDHM-----KQDFFKSAPEGYV----QERTI 99
DB 883 LPKSPFETFFTLKNGKLYLSPCSMLNRYVHQKFTNHQYQELKDPLNYIVETHSENTI 942
QY 100 FFKDDGNYKTR--AEVKFEGDTLVNR-----IELKGIDFKEDGNILGHKLEYNVN 147
DB 943 FFEVDGPYKAMILPSSKEGKGKIKRYAVFNEDGSLAELKGFELKRGEL---QLIKNFQ 999
QY 148 S--HNVYIMAD 156
DB 1000 SDIFKVFLEGD 1010
```

```

RA RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamiyama K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Irie R., Sato H.,
RA Yamanoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
RA Kikuchi H., Murakawa K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA Kawakami B., Ota T., Hayashi K., Sugiyama T., Otsuki T., Ishibashi T.,
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S.,
RA Kawai Y., Wakamatsu A., Kanehori K., Suzuki Y., Sugano S.,
RA Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RT Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RN SEQUENCE OF 336-689 FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=21245130; PubMed=11347906;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 8:85-95(2001).
CC -!- FUNCTION: Converts acetate to acetyl-CoA so that it can be used
CC for oxidation through the tricarboxylic cycle to produce ATP and
CC CO(2) (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + acetate + CoA = AMP + diphosphate +
CC acetyl-CoA.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC Name=2;
CC IsoId=Q9NUB1-1; Sequence=Displayed;
CC IsoId=Q9NUB1-2; Sequence=VSP_007249;
CC Note=No experimental confirmation available;
CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
CC -!- CAUTION: Ref.1 (CAB81884) sequence differs from that shown due to
CC erroneous gene model prediction.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to a
CC frameshift in position 250 and numerous sequencing errors.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL035661; CAB75500.1; -
CC EMBL; AL080312; CAB81884.1; ALT_SEQ.
CC EMBL; BC033261; AAH39261.1; -
CC EMBL; BC044588; AAH44588.1; -
CC EMBL; AK027817; BAB55390.1; ALT_INIT.
CC EMBL; AK022295; BAC03853.1; ALT_SEQ.
CC EMBL; AB058749; BAB47475.1; -
CC Genew: HGNC:16091; ACAS2L.
CC InterPro: IPR000873; AMP-bind.
CC Pfam: PF00501; AMP-binding; 1.
CC PROSITE: PS00455; AMP BINDING; 1.
CC Ligase; Mitochondrion; Transit
KW TRANSIT 1 36 MITOCHONDRION (POTENTIAL).
FT CHAIN 37 689 ACETYL-COENZYME A SYNTHETASE 2-LIKE.
FT DOMAIN 45 53 POLY-ALA.
FT VARSPLIC 446 447 Missing (in isoform 2).
FT CONFLICT 277 277 V -> M (IN REF. 2; AAH39261).
FT CONFLICT 488 488 V -> M (IN REF. 2; AAH44588).
SQ SEQUENCE 689 AA; 74856 MW; 6EB84E39302AD08B CRC64;

Query Match 6.8%; Score 87; DB 1; Length 689;
Best Local Similarity 24.1%; Pred. No. 13;
Matches 33; Conservative 16; Mismatches 36; Indels 52; Gaps 7;

```

```
QY 9 FTGWPIVLVDGNGVHGHKFSVSGEGEGDATYGLTKFKICTTGKLPVWPMTLVTTLSYG 68
Db 473 FFGIVPVLMDKGSV-----VEGNSVSGALCIS-----QAWPGMARTI--- 510
QY 69 VQCFSPYDPHMKQDHFPSKAMPBGYVQERTIPFKDDGNVKTTRA---EVKFGDGLVNRTE 125
Db 511 -----YGDHQRVDYAFKAYP-GY-----YFTGDGAYRTEGGYQYGTGRMDVDVI----- 553
QY 126 LKGIDFKEDGNILGHKL 142
Db 554 -----NISGHR 560

RESULT 8
CP51_CANGA
ID_CP51_CANGA STANDARD; PRT; 533 AA.
AC P50859; Q02312;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 51 (EC 1.14.13.70) (CYP51) (P450-LIA1) (Sterol 14-
alpha-demethylase) (Lanosterol 14-alpha demethylase) (P450-14DM).
GN ERG11 OR CYP51.
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=2001-L5;
RX MEDLINE=96161286; PubMed=8593007;
RA Geber A., Hitchcock C.A., Swartz J.E., Pullen F.S., Marsden K.E.,
RA Kwon-Chung K.J., Bennett J.E.;
RT "Deletion of the Candida glabrata ERG3 and ERG11 genes: effect on cell
RT viability, cell growth, sterol composition, and antifungal
RT susceptibility.";
RL Antimicrob. Agents Chemother. 39:2708-2717(1995).
RN [2]
SEQUENCE OF 60-473 FROM N.A.
RC STRAIN=ATCC 2001;
RX MEDLINE=95081364; PubMed=7989540;
RA Burgener-Kairuz P., Zuber J.P., Jaunin P., Buchman T.G., Bille J.,
RA Rossier M.;
RT "Rapid detection and identification of Candida albicans and
RT Torulopsis (Candida) glabrata in clinical specimens by
RT species-specific nested PCR amplification of a cytochrome P-450
RT lanosterol-alpha-demethylase (Lia1) gene fragment.";
RL J. Clin. Microbiol. 32:1902-1907(1994).
CC -I- FUNCTION: Catalyzes C14-demethylation of lanosterol which is
CC critical for ergosterol biosynthesis. It transforms lanosterol
CC into 4,4'-dimethyl cholesta-8,14,24-triene-3-beta-ol (By
CC similarity).
CC -I- CATALYTIC ACTIVITY: Obtusifolliol + 3 O(2) + 3 NADPH = 4-alpha-
CC methyl-5-alpha-ergosta-8,14,24(28)-trien-3-beta-ol + formate + 3
CC NADP(+) + 3 H(2)O.
CC -I- PATHWAY: Ergosterol biosynthesis.
CC -I- SIMILARITY: Belongs to the cytochrome P450 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L40389; AAB02329.1; -
CC InterPro; IPR001128; Cytochrome_P450.
CC Pfam; PF00067; p450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC Electron transport; Oxidoreductase; Monooxygenase; Membrane; Heme;
KW Sterol biosynthesis; NADP.
FT METAL 472 472 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT CONFLICT 64 64 I -> M (IN REF. 2).
FT CONFLICT 473 473 I -> T (IN REF. 2).
SQ SEQUENCE 533 AA; 61305 MW; A0506C17507E6EF7 CRC64;
Query Match 6.8%; Score 86.5; DB 1; Length 533;
Best Local Similarity 21.4%; Pred. No. 10;
Matches 44; Conservative 32; Mismatches 81; Indels 49; Gaps 8;
QY 25 GHKFSVS---GEGEGDATYGLTKFKICTTGKLPVWPMTLVTTLSYGVQCFSPYDPH--M 79
Db 109 GHEIFNAKLADVSAEAYSHL-----TTPVFGKGVIVDCPNHRLM 149
QY 80 KOHDFFKSAM-PEGYV-----QERTIPFKDDGNVKTTRAEVKFGDGLTVNRTELKGDIF 131
Db 150 EQKKFKVGALTKEAFVRYVPLIAEIIKYFRSKFKINENNSGIVDVNVSQPEM--TIF 207
QY 132 KEDGNILGHKLEYNYNHNVYIMADQKNGIKVKNIRINIEDSGVOLADHQQNTPIGD 191
Db 208 TASRSLGKEMRDKLDTDPAYLYSLDKGFTPINF-VFPNLPLEHYRKRDHAQAIS--- 263
QY 192 GPVLIPDNHYLSTQSALSQDPNEKRD 217
Db 264 -----GTYNMSLIKERREKND 278

RESULT 9
SID_XYLFT
ID_SYL_XYLFT STANDARD; PRT; 879 AA.
AC Q87C65;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR PDI230.
OS Xylella fastidiosa (strain Temeculal / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=183190;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=22421331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T. da Silva F.R.,
RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,
RA Carter H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuranae E.E.,
RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,
RA da Cunha A.F., Penille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sassaki F.T., Sena J.A.D.,
RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,
RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA Kitajima J.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of Xylella
RT fastidiosa.";
RL J. Bacteriol. 185:1018-1026(2003).
CC -I- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

KW Electron transport; Oxidoreductase; Monooxygenase; Membrane; Heme;

```
DR EMBL; AE012557; AAC29080.1; ALT_INIT.
DR HAMAP; MF_00049; -. 1.
DR InterPro; IPR002302; Leu-TRNAsynt1a.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR009008; ValRS_1IERS_edit.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRfams; TIGR00396; leuS_bact; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 45 55 "HIGH" REGION.
FT SITE 637 641 "KMSKS" REGION.
FT BINDING 640 640 ATP (BY SIMILARITY).
SQ SEQUENCE 879 AA; 99823 MW; 4C2BE01B8FDC497E CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 879;
Best Local Similarity 22.3%; Pred. No. 19;
Matches 44; Conservative 28; Mismatches 69; Indels 57; Gaps 10;

QY 50 TTGKLPVWPMTLVTLSYGVCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNY-- 107
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 321 TNEQLPV-WYANFVLMAYGTGAVMVGPHDQDQEF--ANKYGLPIRQVIALKEPKNQDE 377
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 108 -----KTRAEVKFEGDTLVNRIELKGDIFKEDGNILGHKLEYNYSNHNVI 153
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 378 STWEPDVRDWDYADKTR--EFG--LINSAPDGLDYQGAFLAERFE----- 421
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 154 MADKQKNG-IKVNFKIRHNIEDGSVQLADHYQQNTPI-----GDGFVLLPQN 199
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 422 ---RQGRQRVNYRLR---DWGVSQRVWGCPFIVVYCTGCAVPVNPQLPVILPEN 474
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 200 -HYLSTQSALSQDPNEKR 216
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 475 VAFSGTGSPIKTPDEWRK 492
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 10
VIT4_CABEL STANDARD; PRT; 1603 AA.
ID VIT4_CABEL
AC P18947; Q9BPP3;
DT 01-NOV-1990 (Rel. 16, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vitellogenin 4 precursor.
GN Vit-4 OR F59D8.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2.
RA Waterston R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-282 FROM N.A.
RA Blumenthal T., Spieth J., Zucker E.;
RL Submitted (NOV-1986) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-71 FROM N.A.
RX MEDLINE=85269643; PubMed=4022780;
RA Spieth J., Denison K., Kirtland S., Cane J., Blumenthal T.;
RT "The C. elegans vitellogenin genes: short sequence repeats in the
RT promoter regions and homology to the vertebrate genes.";
RL Nucleic Acids Res. 13:5283-5295(1985).
CC -!- FUNCTION: Precursor of the egg-yolk proteins that are sources of
CC nutrients during embryonic development (Potential).
CC -!- SUBCELLULAR LOCATION: Secreted
CC CC
CC -!- TISSUE SPECIFICITY: Synthesized in Caenorhabditis only by 32 cells
CC including the intestine of adult hermaphroditic individuals; they
CC are cotranslationally secreted into the body cavity and
```

```
CC subsequently taken up by the gonad.
CC -!- SIMILARITY: Contains 1 VWFD domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AC024137; AAK09074.1; -.
DR EMBL; M11498; AAA28163.1; -.
DR EMBL; X02754; CAA36531.1; -.
DR FIR; A43084; A43084.
DR WormPep; F59D8.2; CE26817.
DR InterPro; IPR001747; Lipid transprt_N.
DR InterPro; IPR001846; VWF D.
DR Pfam; PF01347; Vitellogenin_N; 1.
DR SMART; SM00638; LPD_N; 1.
DR SMART; SM00216; VWF; 1.
KW Storage protein; Multigene family; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 1603
FT DOMAIN 1308 1455 VWFD.
FT CONFLICT 30 30 Y -> V (IN REF. 3).
FT CONFLICT 169 169 L -> V (IN REF. 2).
FT CONFLICT 183 187 EVAYT -> RSLRH (IN REF. 2).
FT CONFLICT 275 275 T -> S (IN REF. 2).
SQ SEQUENCE 1603 AA; 186307 MW; E30170325BC99BB CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 1603;
Best Local Similarity 23.4%; Pred. No. 38;
Matches 52; Conservative 32; Mismatches 69; Indels 69; Gaps 12;

QY 1 MYSKGEEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGLTLKFLICTTGKLPVPWPT 60
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 162 MESDKSLFFNVHEKTMGDCV---AYTVQEG-GKTIYTKSVNPKCITR-----PE 211
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEG-YVOERTIF---FKDDG----- 105
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 212 TAYGLRFGSEC-----KECEKGQFVQPQTIVTYTFRKNEKLQSEVNSIYT 257
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 106 -----NYKTRAEVKFEGDTLVNRIELKGDIFKEDGNILGHKLEYNYSNHNVIAD 156
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 258 LNVNGQEVVKSETRAKVFVEESKINR-EIK-----KVSQPKKEIYVSMENKLIQ 308
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 157 KQKNG-----IKVNFKIRHNIEDGSVQLADHYQQNTTP 188
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 309 FYKQGDKAENVNPFKAIEIQKV-EQLEEIFRQIQEH-EQNTTP 348
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 11
TRNA_CAMJE STANDARD; PRT; 357 AA.
ID TRNA_CAMJE
AC Q9P932;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE tRNA (Uracil-5)-methyltransferase (EC 2.1.1.35) (cRNA (M-5-Us4) -
DE methyltransferase) (RUMT).
GN TRNA OR CU0831C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagsels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
```



RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
 RA Whitehead S., Barrell B.G.;  
 RT "The genome sequence of the food-borne pathogen *Campylobacter jejuni*  
 RL reveals hypervariable sequences.";   
 RT Nature 403:665-668(2000).  
 CC -!- FUNCTION: Catalyzes the formation of 5-methyl-uridine at position  
 CC 54 (W-5-U54) in all tRNA (By similarity).  
 CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-  
 CC homocysteine + tRNA containing thymine.  
 CC -!- SIMILARITY: Belongs to the RNA M5U methyltransferase family. TrmA  
 CC subfamily.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AL139076; CAB73096.1; -;  
 DR PIR; G81355; G81355.  
 DR HAMAP; MF\_01011; -; 1.  
 DR InterPro; IPR000051; SAM\_bind.  
 DR InterPro; IPR001566; TrmA.  
 DR PROSITE; PS01230; TRMA 1; 1.  
 DR PROSITE; PS01231; TRMA 2; FALSE NEG.  
 KW Transferase; Methyltransferase; tRNA processing; Complete proteome.  
 FT DOMAIN 207 213 S-ADENOSYLMETHIONINE BINDING (BY  
 FT ACT\_SITE 315 315 BY SIMILARITY.  
 FT ACT\_SITE 357 357 BY SIMILARITY.  
 SQ SEQUENCE 357 AA; 42276 MW; CEC528347CEB497 CRC64;  
 Query Match 6.8%; Score 86; DB 1; Length 357;  
 Best Local Similarity 24.8%; Pred. No. 7;  
 Matches 30; Conservative 18; Mismatches 39; Indels 34; Gaps 5;  
 QY 80 KOHDFPKSAMPEGYVOERTIFFKDGNYKTRAEVKE--EGDTLV-----NRIELKG 128  
 DB 14 EKHSEFKIKYFKFYTKDFKLFASKDGHYTRAEISFYHENDILFYAMFPDKSKKVIIEY 73  
 QY 129 IDPKED-----GNILGHKLEYNNSHNVMADKQKNGIKVNFKIRHNIE 173  
 DB 74 LQFADEKICAFMPLLELYLRQDNKLEKL-----FGVEFLTKQE--LSITLLYHKNIE 125  
 QY 174 D 174  
 DB 126 D 126  
 RESULT 12  
 PSBC\_CYAPA STANDARD; PRT; 461 AA.  
 AC P48104;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Photosystem II 44 kDa reaction center protein (P6 protein) (CP43).  
 GN PSBC.  
 OS Cyanophora paradoxa.  
 OG Cyanelle.  
 OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.  
 OX NCBI\_TaxID=2762;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=UTEX LB 555 / Pringsheim;  
 RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,  
 RA Bryant D.A.;  
 RT "Nucleotide sequence of the cyanelle DNA from *Cyanophora paradoxa*.";   
 RL Plant Mol. Biol. Rep. 13:327-332(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=UTEX LB 555 / Pringsheim;

RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,  
 RA Farley J.Y., Schlachter W.M., Chung S., Newmann-Spallart C.,  
 RA Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;  
 RT "The complete sequence of the cyanelle genome of *Cyanophora paradoxa*:  
 RT the genetic complexity of a primitive plastid.";   
 RL (in) Schenk H.E.A., Hertmann R., Jeon K.W., Mueller N.E.,  
 RL Schwemmler W. (eds.);  
 RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg  
 RL (1997).  
 CC -!- FUNCTION: The 43 kDa protein (p6) is a component of the core of  
 CC photosystem II. It is a chlorophyll binding protein.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Cyanelle  
 CC thylakoid membrane.  
 CC -!- SIMILARITY: Belongs to the psbB / psbC family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; U30821; AAA81279.1; -;  
 DR PIR; T06936; T06936.  
 DR InterPro; IPR005869; Photo44.  
 DR InterPro; IPR000932; PSIIProt.  
 DR Pfam; PF00421; PSII; 1.  
 DR TIGRfam; TIGR01153; psbC; 1.  
 KW Photosynthesis; Photosystem II; Thylakoid; Chlorophyll; Cyanelle;  
 KW Transmembrane.  
 SQ SEQUENCE 461 AA; 50402 MW; 62664E239E13B9C1 CRC64;  
 Query Match 6.8%; Score 86; DB 1; Length 461;  
 Best Local Similarity 21.0%; Pred. No. 9.4;  
 Matches 57; Conservative 35; Mismatches 89; Indels 90; Gaps 13;  
 QY 25 GHKFSVSGEGEDATYVKLTIKFICTGKL-----PVPWP--TLVTT---LSYGV--- 69  
 DB 208 GGNWISVDNMEDIGGHIWLAFLIIGVWHILYKFPFWARRALVWSEAYLSYLAAL 267  
 QY 70 -----QCFSYPDHMKQHDFFKSGAMPE-GYVQERTIFFKDD-----GN 106  
 DB 268 ALMGFIANCFFVFNNTAYPSEFFGTPGPASQAQAFTLVLRDQLGANVSGAQGTGLGK 327  
 QY 107 YKTRA---EVKPEGDTLVNRIELKGIDPKEDGNILGHKLEYNNSHNVMADKQKNGIK 163  
 DB 328 YLMRSPSGEIIIFGGETM-----RFDTRAPWLEPLRGANGLD 364  
 QY 164 VNFKIRHNIEDGSLADHYQONTPIGD---GPVLLPDN---HYLSTQSALSKDPNEKRD 217  
 DB 365 LT-KIYDIQPWQERRAAEYMTWAPLGLSLNSVGGVATEINSVNVSPRSWLS-----TS 417  
 QY 218 HNVLAGFV-----TAAGITLGM 235  
 DB 418 HPVLGFFELFIGHLHAGRAAASGGFEKGLD 448  
 RESULT 13  
 YC03\_KLEPN STANDARD; PRT; 504 AA.  
 ID YC03\_KLEPN  
 AC Q48449;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 55.8 kDa protein in cps region (ORF3).  
 OS Klebsiella pneumoniae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Klebsiella.  
 OX NCBI\_TaxID=573;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Chedid;

```

RX MEDLINE=95204345; PubMed=7896702;
RA Arakawa Y., Wacharotayankun R., Nagatsuka T., Ito H., Kato N.,
RT "Genomic organization of the Klebsiella pneumoniae cps region
RT responsible for serotype K2 capsular polysaccharide synthesis in the
RT virulent strain Chediak."
RL J. Bacteriol. 177:1788-1796(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D21242; BAA04774.1; -.
KW Hypothetical protein.
SQ SEQUENCE 504 AA; 55782 MW; AD887595CFDFDAB CRC64;

Query Match      6.8%; Score 86; DB 1; Length 504;
Best Local Similarity 24.0%; Pred. No. 10;
Matches 46; Conservative 27; Mismatches 71; Indels 48; Gaps 9;

QY 70 QCFSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNKYKTRAEVKPEGD-TLVNRTELKG 128
DB 293 QSFSSFWDFGTGHD-----NTGTDNEPGNLQAGDFPKLPTLGNPVSFYG 339

QY 129 IDPKED-----GNILGHKLEYNNSHNVYIMA-DKQNGIKVNFKIRHNI-BD 174
DB 340 QMWGEDESGLPSANFLGEGIEGHH-GWGKDAVNWYVEAHDTRTNMSRTNYSYTHIYKD 398

QY 175 GSVQLADHYQONTPIGG-----PVLPPNHVLSQSALSK-DPNEKRDHVMV 221
DB 399 G-----YQQQYPLGDAMGGDQLFAGKVELITENNQRWSTRLAYAKVNPXQDSINKAP 452

QY 222 AGFVTAAGITLG 233
DB 453 PHSDTLKGVLG 464

RESULT 14
ID AMPA WIGBR STANDARD; PRT; 501 AA.
AC Q8D235;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Probable cytosol aminopeptidase (EC 3.4.11.1) (Leucine aminopeptidase)
DE (LAP) (Leucyl aminopeptidase).
GN PEPA OR WIGBR4590
OS Wigglesworthia glossinidia brevipalpis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Wigglesworthia.
OX NCBI_TaxID=36870;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22297718; PubMed=12219091;
RA Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,
RA Aksoy S.;
RT flies, Wigglesworthia glossinidia."
RL Nat. Genet. 32:402-407(2002).
CC -!- FUNCTION: Presumably involved in the processing and regular
CC turnover of intracellular proteins. Catalyzes the removal of
CC unsubstituted N-terminal amino acids from various peptides (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: Release of an N-terminal amino acid, Xaa-|-
CC Xbb-, in which Xaa is preferably Leu, but may be other amino acids
CC including Pro although not Arg or Lys, and Xbb may be Pro.
CC -!- COFACTOR: Binds 2 manganese ions per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family M17.

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB063522; BAC24605.1; -.
DR HAMAP; MF_00181; -.
DR InterPro; IPR000819; Peptidase M17 C.
DR InterPro; IPR008283; Peptidase M17_N.
DR Pfam; PF00883; Peptidase M17; 1.
DR Pfam; PF02789; Peptidase M17_N; 1.
DR PRINTS; PR00481; LAMOPPTDASE.
DR PROSITE; PS00631; CYTOSOL_AP; 1.
KW Hydrolase; Aminopeptidase; Manganese; Complete proteome.
FT ACT_SITE 282..282 POTENTIAL.
FT ACT_SITE 356..356
FT METAL 270..270 MANGANESE 2 (BY SIMILARITY).
FT METAL 275..275 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 293..293 MANGANESE 2 (BY SIMILARITY).
FT METAL 352..352 MANGANESE 1 (BY SIMILARITY).
FT METAL 354..354 MANGANESE 1 AND 2 (BY SIMILARITY).
SQ SEQUENCE 501 AA; 56643 MW; 4E181EBEB481FE3 CRC64;

Query Match      6.7%; Score 85.5; DB 1; Length 501;
Best Local Similarity 21.5%; Pred. No. 11;
Matches 35; Conservative 24; Mismatches 81; Indels 23; Gaps 5;

QY 93 YVOERTIPKDDGNKYKTRAEVKPEGDTLVNRTELKGIDF-KEDGNILGHKLEYNNSHNV 151
DB 153 YKKEIMFYIHDENEIK-QANIAISHSVIS-----KGIIITKLGWNPSCFDPHYLSHOS 207

QY 152 YIMADQKNGIKVNFKIRHNIEDGVSQVLADH-----YQONTPIGDPVLL 196
DB 208 YILDKYSEKISVEIMDHKKIKNGWNYLVHVSCKSNPYLSIIKYNENKFGKSPILL 267

QY 197 PDNHVLSQSALSKDPNEKRDHVM--LAGFVTAAGITLGMDEL 237
DB 268 ICKGLTFDGGISIKPSNNMDEMCKFDMCGAAAVLGMVHAISEL 310

RESULT 15
ID DPOL_HPBHE STANDARD; PRT; 788 AA.
AC P13846;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE P protein [Includes: DNA-directed DNA polymerase (EC 2.7.7.7); RNA-
DE directed DNA polymerase (EC 2.7.7.49); Ribonuclease H (EC 3.1.26.4)].
GN P.
OS Heron hepatitis B virus.
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=28300;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88333160; PubMed=3418788;
RA Sprengel R., Kaleta E.F., Will H.;
RT "Isolation and characterization of a hepatitis B virus endemic in
RT herons."
RL J. Virol. 62:3832-3839(1988).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA}(N).
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
CC phosphomonoester.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

Search completed: June 21, 2004, 15:55:24  
Job time : 7.55556 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:45:30 ; Search time 30.6667 Seconds  
(without alignments)  
2458.984 Million cell updates/sec

Title: US-09-887-784-222a

Perfect score: 1274

Sequence: 1 MVSKGELFTGVVPIVLVD.....VLagFVTRAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1238	97.2	238	2 Q8GHE2	Q8ghe2 azotobacter
2	1235	96.9	238	5 Q93125	Q93125 aequorea vi
3	1233	96.8	238	2 Q8GHE4	Q8ghe4 azomonas ag
4	1232	96.7	238	2 Q8GHE3	Q8ghe3 azotobacter
5	1200	94.2	238	5 Q17105	Q17105 aequorea vi
6	1185	93.0	238	5 Q17106	Q17106 aequorea vi
7	1080	84.8	238	5 Q8WTC6	Q8wtc6 aequorea ma
8	1076	84.5	238	5 Q8WP95	Q8wp95 aequorea ma
9	1072	84.1	238	5 Q8WTC4	Q8wtc4 aequorea ma
10	1070	84.0	238	5 Q8WTD0	Q8wtcd0 aequorea ma
11	1069	83.9	238	5 Q8WTC8	Q8wtc8 aequorea ma
12	1069	83.9	238	5 Q8WTC9	Q8wtc9 aequorea ma
13	1067	83.8	238	5 Q8WTC7	Q8wtc7 aequorea ma
14	1065	83.6	238	5 Q8WTC5	Q8wtc5 aequorea ma
15	252.5	19.8	225	5 Q95UA7	Q95ua7 montastraea
16	252.5	19.8	225	5 Q720W5	Q720w5 montastraea

17	247	19.4	225	5	Q963F5	Q963f5 montastraea
18	244.5	19.2	236	5	Q8T6U0	Q8t6u0 dendroneph
19	242.5	19.0	225	5	Q720W9	Q720w9 montastraea
20	240	18.8	225	5	Q8I6J8	Q8i6j8 trachyphyl
21	238.5	18.7	266	5	Q9U6Y3	Q9u6y3 clavularia
22	233	18.3	225	5	Q720W4	Q720w4 montastraea
23	232	18.2	224	5	Q8MU48	Q8mu48 montastraea
24	232	18.2	225	5	Q8T5F1	Q8t5f1 montastraea
25	214	16.8	227	5	Q720W6	Q720w6 montastraea
26	214	16.8	234	5	Q720W7	Q720w7 montastraea
27	212.5	16.7	259	5	Q8MMA2	Q8mma2 agaricia fr
28	210	16.5	239	5	Q8MMA1	Q8mma1 agaricia ag
29	209	16.4	234	5	Q8T5F2	Q8t5f2 montastraea
30	209	16.4	234	5	Q8MU47	Q8mu47 montastraea
31	208.5	16.4	229	5	Q9U6Y6	Q9u6y6 anemonia ma
32	206	16.2	227	5	Q962P9	Q962p9 montastraea
33	206	16.2	227	5	Q720W8	Q720w8 montastraea
34	205.5	16.1	232	5	Q9GFI5	Q9gfi5 anemonia su
35	204.5	16.1	214	5	Q86LV7	Q86lv7 meandrina m
36	204	16.0	221	5	Q95P04	Q95p04 gonopora t
37	203.5	16.0	214	5	Q86LV8	Q86lv8 meandrina m
38	203.5	16.0	238	5	Q9BLY9	Q9bly9 renilla mue
39	202	15.9	227	5	Q95V70	Q95vt0 montastraea
40	201.5	15.8	225	5	Q9U6Y8	Q9u6y8 discosoma s
41	201.5	15.8	232	5	Q9GZ28	Q9gz28 anemonia su
42	199	15.6	235	5	Q8T5F0	Q8t5f0 scolymia cu
43	198.5	15.6	222	5	Q72168	Q72168 cerianthus
44	198.5	15.6	225	5	Q8T6T9	Q8t6t9 radianthus
45	197.5	15.5	232	5	Q9U6Y7	Q9u6y7 discosoma s

ALIGNMENTS

RESULT 1

Q8GHE2	PRELIMINARY;	PRT;	238 AA.
ID	Q8GHE2		
AC	Q8GHE2;		
DT	01-WAR-2003 (TrEMBLrel. 23, Created)		
DT	01-WAR-2003 (TrEMBLrel. 23, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Green fluorescence protein.		
DE	2289GFP.		
GN	Azotobacter vinelandii.		
OS	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;		
OC	Pseudomonadaceae; Azotobacter.		
OX	NCBI_TaxID=354;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=DSM2289;		
RA	Koranyi P., Berenyi M., Burg K.;		
RT	"Occurrence of green fluorescence protein in diazotrophic bacteria		
RL	Azomonas and Azotobacter."		
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF324408; AAN86140.1; -.		
DR	GO; GO:0006091; P:energy pathways; IEA.		
DR	InterPro; IPR000786; GFP_like.		
DR	InterPro; IPR000786; Green_fl_protein.		
DR	Pfam; PF01353; GFP; 1.		
DR	PRINTS; PD01229; GFLUORESCENT.		
DR	ProDom; PD013756; Green_fl_protein; 1.		
SQ	SEQUENCE 238 AA; 26914 MW; F84840F1F9064018 CRC64;		

Query Match	97.2%;	Score 1238;	DB 2;	Length 238;
Best Local Similarity	97.5%;	Pred. No. 1.1e-95;		
Matches 232;	Conservative	2;	Mismatches 4;	Indels 0;
Gaps	0;			
QY	2	VS	KGELFTGVVPIVLVDGVNGHKFSVSGEGDATYKGLTKFKICTTGKLPVWP	61
Db	1	MS	KGELFTGVVPIVLVDGVNGHKFSVSGEGDATYKGLTKFKICTTGKLPVWP	60
QY	62	VT	LSYGVQCFRYPDPMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKEGDTLV	121

Db 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120  
QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLAD 181  
Db 121 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLAD 180  
QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMVLAVFTAAAGITLGMDELYK 239  
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMVLAVFTAAAGITLGMDELYK 238

RESULT 2

Q93125 ID Q93125 PRELIMINARY; PRT; 238 AA.  
AC Q93125  
DT 01-FEB-1997 (TrEMBLrel. 02, Created)  
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Green fluorescent protein mutant 3.  
GN GFP.  
OS Aequorea victoria (Jellyfish).  
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;  
OC Aequoreidae; Aequorea.  
OX NCBI\_TaxID=6100;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96305137; PubMed=8707053;  
RA Cormack B.P., Valdivia R.H., Falkow S.;  
RT "FACS-optimized mutants of the green fluorescent protein (GFP).";  
RL Gene 173:33-38(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Cormack B.P., Bertram G., Egerton M., Gow N.A.R., Falkow S.,  
RA Brown A.J.P.;  
RT expression in *Candida albicans*.  
RL Microbiology 0:0-0(1996).  
DR EMBL; U73901; AAB18957.1; -.  
DR HSSP; P42212; 1BFP.  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR009017; GFP\_like.  
DR InterPro; IPR000786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFP; 1.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
SQ SEQUENCE 238 AA; 26840 MW; A28622809A9DEA60 CRC64;

Query Match 96.9%; Score 1235; DB 5; Length 238;  
Best Local Similarity 97.1%; Pred. No. 1.9e-95;  
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
QY 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKLFCITCTGKLPVWPPTL 61  
Db 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKLFCITCTGKLPVWPPTL 60  
QY 62 VTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121  
Db 61 VTTFTGTVQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120  
QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLAD 181  
Db 121 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLAD 180  
QY 192 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMVLAVFTAAAGITLGMDELYK 239  
Db 191 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMVLAVFTAAAGITLGMDELYK 238

RESULT 3

Q8GHE4 ID Q8GHE4 PRELIMINARY; PRT; 238 AA.  
AC Q8GHE4  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Green fluorescence protein.  
GN 375GFP.  
OS Azomonas agilis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Azomonas.  
OX NCBI\_TaxID=116849;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Koranyi P., Berenyi M., Burg K.;  
RT "Occurrence of green fluorescence protein in diazotrophic bacteria  
RT Azomonas and Azotobacter.";  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF324405; AAN86137.1; -.  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR009017; GFP\_like.  
DR InterPro; IPR000786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFP; 1.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
SQ SEQUENCE 238 AA; 26902 MW; 15FE9B9C5B4F6B89 CRC64;  
Query Match 96.8%; Score 1233; DB 2; Length 238;  
Best Local Similarity 97.1%; Pred. No. 2.9e-95;  
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
QY 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKLFCITCTGKLPVWPPTL 61  
Db 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKLFCITCTGKLPVWPPTL 60  
QY 62 VTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121  
Db 61 VTTFTGTVQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120  
QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLAD 181  
Db 121 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLAD 180  
QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMVLAVFTAAAGITLGMDELYK 239  
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMVLAVFTAAAGITLGMDELYK 238

RESULT 4

Q8GHE3 ID Q8GHE3 PRELIMINARY; PRT; 238 AA.  
AC Q8GHE3  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DE Green fluorescence protein.  
GN 85GFP.  
OS Azotobacter vinelandii.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Azotobacter.  
OX NCBI\_TaxID=354;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Koranyi P., Berenyi M., Burg K.;  
RT "Occurrence of green fluorescence protein in diazotrophic bacteria  
RT Azomonas and Azotobacter.";  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF324406; AAN86138.1; -.  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR009017; GFP\_like.  
DR InterPro; IPR000786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFP; 1.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
SQ SEQUENCE 238 AA; 26887 MW; E0E1616BD2AF6188 CRC64;  
Query Match 96.7%; Score 1232; DB 2; Length 238;

```
Best Local Similarity 97.1%; Pred. No. 3.5e-95;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVWPILVELDGVNGHKFVSVEGEGDATYKGLTLKFCITCTGKLPVWPPTL 61
DB 1 MSKGEELFTGVWPILVELDGVNGHKFVSVEGEGDATYKGLTLTFCITCTGKLPVWPPTL 60

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB 61 VTTFSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKVNFKIRNIHEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKVNFKIRNIHEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLGAVTAAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLGAVTAAAGITLGMDELYK 238

RESULT 5
Q17105
ID Q17105 PRELIMINARY; PRT; 238 AA.
AC Q17105;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83959; CAA58789.1; -.
DR PIR; JS0692; J01514.
DR HSSP; P42212; 1GFL.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748E44 CRC64;

Query Match 94.2%; Score 1200; DB 5; Length 238;
Best Local Similarity 93.7%; Pred. No. 1.7e-92;
Matches 223; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVWPILVELDGVNGHKFVSVEGEGDATYKGLTLKFCITCTGKLPVWPPTL 61
DB 1 MSKGEELFTGVWPILVELDGVNGHKFVSVEGEGDATYKGLTLNFCITCTGKLPVWPPTL 60

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB 61 VTTFSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKVNFKIRNIHEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKVNFKIRNIHEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLGAVTAAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLGAVTAAAGITLGMDELYK 238

RESULT 6
Q17106
ID Q17106 PRELIMINARY; PRT; 238 AA.
```

```
Q17106;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83960; CAA58790.1; -.
DR PIR; JS0692; J01514.
DR HSSP; P42212; 1BFF.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; BD4648262D8EABD4 CRC64;

Query Match 93.0%; Score 1185; DB 5; Length 238;
Best Local Similarity 92.9%; Pred. No. 3e-91;
Matches 221; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVWPILVELDGVNGHKFVSVEGEGDATYKGLTLKFCITCTGKLPVWPPTL 61
DB 1 MSKGEELFTGVWPILVELDGVNGHKFVSVEGEGDATYKGLTLKFCITCTGKLPVWPPTL 60

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB 61 VTTFSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKVNFKIRNIHEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKVNFKIRNIHEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLGAVTAAAGITLGMDELYK 239
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLGAVTAAAGITLGMDELYK 238

RESULT 7
Q8WTC6
ID Q8WTC6 PRELIMINARY; PRT; 238 AA.
AC Q8WTC6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GFPXT19uv;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435431; RAL33916.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP_like.
DR Pfam; PF01353; GFP; 1.
```





```
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435427; AAL33912.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26997 MW; 5F80A192173CB84D CRC64;

Query Match      84.0%; Score 1070; DB 5; Length 238;
Best Local Similarity 81.1%; Pred. No. 1.5e-81;
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSFGEGDGYGKLTKEICTTGKLPVWPPTL 61
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MSKGEELFTGIVPVLIELDGDVHGKFSVSGEGDADYGKLEIKETCTGKLPVWPPTL 60

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 VTTLGVIQCFARYPEHMKMNDFFKSAMPEGYIERTIFFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 NRIELKGMDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLAVGFTVTAAGITLGMDELYK 239
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 HYQTNVPLGDGPVLLPINHYLSQTAISKDRNETRDMHVFLEFFSACGHTGMDLYK 238

RESULT 11
Q8WTC8 PRELIMINARY; PRT; 238 AA.
AC Q8WTC8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RC STRAIN=GFPxm163;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435428; AAL33913.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27031 MW; 5F80A19C19DC584D CRC64;

Query Match      83.9%; Score 1069; DB 5; Length 238;
Best Local Similarity 81.1%; Pred. No. 1.5e-81;
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSFGEGDGYGKLTKEICTTGKLPVWPPTL 61
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MSKGEELFTGIVPVLIELDGDVHGKFSVSGEGDADYGKLEIKETCTGKLPVWPPTL 60

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 VTTLGVIQCFARYPEHMKMNDFFKSAMPEGYIERTIFFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 NRIELKGMDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLAVGFTVTAAGITLGMDELYK 239
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 HYQTNVPLGDGPVLLPINHYLSQTAISKDRNETRDMHVFLEFFSACGHTGMDLYK 238

RESULT 13
Q8WTC7 PRELIMINARY; PRT; 238 AA.
AC Q8WTC7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435429; AAL33914.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;

Query Match      83.9%; Score 1069; DB 5; Length 238;
Best Local Similarity 81.1%; Pred. No. 1.5e-81;
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSFGEGDGYGKLTKEICTTGKLPVWPPTL 61
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MSKGEELFTGIVPVLIELDGDVHGKFSVSGEGDADYGKLEIKETCTGKLPVWPPTL 60

QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 VTTLGVIQCFARYPEHMKMNDFFKSAMPEGYIERTIFFQDDGKYKTRGEVKFEGDTLV 120
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:41:49 ; Search time 47.1111 Seconds  
(without alignments)  
1433.395 Million cell updates/sec

Title: US-09-887-784-222G

Perfect score: 1276

Sequence: 1 MYSKGEELFTGVVILVELD.....VLGFFVTRAGITLGMDELK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: Genesecp1980s:\*
- 2: Genesecp1990s:\*
- 3: Genesecp2000s:\*
- 4: Genesecp2001s:\*
- 5: Genesecp2002s:\*
- 6: Genesecp2003as:\*
- 7: Genesecp2003bs:\*
- 8: Genesecp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1266	99.2	239	5	AAE17518 Enhanced
2	1266	99.2	363	6	ABR40352 Human ami
3	1266	99.2	893	4	AAG65781 Amino aci
4	1266	99.2	1132	4	AAG65782 Amino aci
5	1258	98.6	239	5	AAE17517 Enhanced
6	1255	98.4	239	3	AAE17517 Enhanced
7	1255	98.4	239	3	AAE17517 Enhanced
8	1255	98.4	239	3	AAE17517 Enhanced
9	1255	98.4	239	4	AAE17517 Enhanced
10	1255	98.4	239	4	AAE17517 Enhanced
11	1255	98.4	239	4	AAE17517 Enhanced
12	1255	98.4	239	5	AAE17517 Enhanced
13	1255	98.4	239	5	AAE17517 Enhanced
14	1255	98.4	239	5	AAE17517 Enhanced
15	1255	98.4	239	6	AAE17517 Enhanced
16	1255	98.4	239	6	AAE17517 Enhanced
17	1255	98.4	239	6	AAE17517 Enhanced
18	1255	98.4	239	6	AAE17517 Enhanced
19	1255	98.4	239	7	AAE17517 Enhanced
20	1255	98.4	239	7	AAE17517 Enhanced
21	1255	98.4	239	7	AAE17517 Enhanced
22	1255	98.4	239	7	AAE17517 Enhanced
23	1255	98.4	239	7	AAE17517 Enhanced
24	1255	98.4	239	7	AAE17517 Enhanced
25	1255	98.4	239	7	AAE17517 Enhanced

26	1255	98.4	265	2	AAW97451	Aaw97451 Wild-type
27	1255	98.4	268	5	AAU99803	Aau99803 Biomembra
28	1255	98.4	270	5	AAU99802	Aau99802 Biomembra
29	1255	98.4	272	5	AAU99800	Aau99800 Biomembra
30	1255	98.4	273	5	AAU99801	Aau99801 Biomembra
31	1255	98.4	280	5	AAU99807	Aau99807 Biomembra
32	1255	98.4	281	3	AAU50142	Aay50142 Green flu
33	1255	98.4	281	3	AAE24252	Aab24252 EGFP-MODC
34	1255	98.4	281	5	AAU10888	Aau10888 EGFP-MODC
35	1255	98.4	286	7	ADE28562	Ade28562 EGFP/ hum
36	1255	98.4	289	7	ADE28564	Ade28564 EGFP/ hum
37	1255	98.4	290	7	ADE28568	Ade28568 EGFP/ hum
38	1255	98.4	290	7	ADE28566	Ade28566 EGFP/ hum
39	1255	98.4	294	3	AAE22860	Aab22860 GFP-DEVD-
40	1255	98.4	294	3	AAE22860	Aab22860 GFP-DEVD-
41	1255	98.4	294	5	ABG94422	Abg94422 Recombina
42	1255	98.4	308	2	AAE42181	Aay42181 EGFP/DRM
43	1255	98.4	320	6	ABR83620	Abr83620 HUB1-GFP
44	1255	98.4	323	3	AAE54359	Aay54359 GFP mutan
45	1255	98.4	323	6	ABR83621	Abr83621 RUB1-GFP

ALIGNMENTS

RESULT 1

AAE17518

ID AAE17518 standard; protein; 239 AA.

AC AAE17518;

DT 22-APR-2002 (first entry)

DE Enhanced F64L-E222G jellyfish green fluorescent protein mutant.

KW Jellyfish; green fluorescent protein; GFP; protein redistribution;  
cellular function; genetic reporter; mutant; Stoke's shift; mutein.

OS Aequorea victoria.

OS Synthetic.

FH Key Location/Qualifiers

FT Misc-difference 65

FT /note= "Wild type Phe substituted with Leu; This  
corresponds to position 64 in the wild type protein"

FT Misc-difference 223

FT /note= "Wild type Glu substituted with Gly; This  
corresponds to position 222 in the wild type protein"

FT WO200198338-A2.

PD 27-DEC-2001.

PF 18-JUN-2001; 2001WO-EP006849.

PR 19-JUN-2000; 2000DK-00000953.

PR 20-JUN-2000; 2000US-0212681P.

PR 10-MAY-2001; 2001DK-00000739.

PR 10-MAY-2001; 2001US-0290170P.

XX (BIOI-) BIOIMAGE AS.

PI Bjorn SP, Pagliaro L, Thastrup O;

XX WPI; 2002-098224/13.

DR N-PSDB; AAD28163.

PT Novel fluorescent protein in in vitro assay for measuring protein kinase  
activity or dephosphorylation activity, or for measuring protein  
redistribution, has a green fluorescent protein with F64L and E222G  
mutation.

XX Claim 9; Page 37; 41pp; English.

PS

XX The invention relates to a fluorescent protein derived from green  
CC fluorescent protein (GFP) or its analogue. The GFP containing mutations  
CC at F64L and E222G has a bigger compared to other GFP's making it very  
CC suitable for high throughput screening due to better resolution. The  
CC fluorescent protein is useful in vitro assays for measuring protein  
CC kinase activity or dephosphorylation activity, or for measuring cellular  
CC redistribution. The fluorescent protein is useful in studying cellular  
CC functions in living cells; as protein tags in transgenic animals, living  
CC and fixed cells; organelle tags, secretion marker and genetic reporter.  
CC The fluorescent protein is also useful as a cell or organelle integrity  
CC marker, a marker for changes in cell morphology, as transfection marker,  
CC and as a marker to be used in combination with fluorescence activated  
CC cell sorting (FACS). The novel proteins can also be used as reporters to  
CC monitor live or dead biomass of organisms, such as fungi. The fluorescent  
CC protein is also useful as markers in transcriptional and translational  
CC fusions for performing transposon vector mutagenesis and as a reporter  
CC for bacterial detection. Transposons encoding the fluorescent protein are  
CC useful for screening promoters and for tagging plasmids and chromosomes.  
CC The fluorescent protein engineered into the genome of a phage is useful  
CC for designing diagnostic tool. The present sequence is a DNA encoding  
CC enhanced F64L-E222G jellyfish green fluorescent protein (GFP) mutant  
XX  
SQ Sequence 239 AA;

Query Match 99.2%; Score 1266; DB 5; Length 239;  
Best Local Similarity 99.8%; Pred. No. 1.4e-122;  
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDGYGKLTAKFICTTGKLPVWPWT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDGYGKLTAKFICTTGKLPVWPWT 60  
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVLGGFVTAAGITLGMDELYK 239  
DB 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVLGGFVTAAGITLGMDELYK 239

RESULT 2  
ID ABR40352  
XX ABR40352 standard; protein; 363 AA.  
AC ABR40352;

XX 08-JUL-2003 (first entry)  
DT Human amino acid sequence SEQ ID NO: 6.  
DE Human; heterologous conjugate; intracellular protein.

XX Homo sapiens.  
OS Aequoria victoria.  
XX WO2003029827-A2.  
XX 10-APR-2003.  
PD 01-OCT-2002; 2002WO-DK000651.  
PF 01-OCT-2001; 2001DK-00001433.  
PR 11-OCT-2001; 2001US-0328996P.

XX (BIOL-) BIOMAGE AS.  
PA Terry BR, Nielsen SJ;  
PI

XX WPI: 2003-430211/40.  
DR N-PSDB; ACC72604.

XX Novel cell for identifying modulators of protein interaction, contains a  
PT first conjugate comprising anchor protein, second conjugate having type B  
PT interactor protein and third conjugate with detectable group.

XX Disclosure; Page 112-113; 118pp; English.

XX The invention relates to a novel cell, comprising three heterologous  
CC conjugates (HC), a first HC (HC1) comprising an anchor protein that  
CC specifically binds to an internal structure within the cell conjugated to  
CC an interactor protein (IP) of type A, a second HC (HC2) comprising IP of  
CC type B conjugated to a first protein of interest, and a third HC (HC3)  
CC comprising a second protein of interest conjugated to detectable group.  
CC The cell is useful for detecting if a compound disrupts or induces the  
CC interaction between two intracellular proteins. The cell is also useful  
CC for screening compounds that modulate the interaction between two  
CC intracellular proteins. The present sequence is used in the  
CC exemplification of the invention

XX Sequence 363 AA;

Query Match 99.2%; Score 1266; DB 6; Length 363;  
Best Local Similarity 99.6%; Pred. No. 2.6e-122;  
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDGYGKLTAKFICTTGKLPVWPWT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDGYGKLTAKFICTTGKLPVWPWT 60  
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVLGGFVTAAGITLGMDELYK 239  
DB 181 DHYQNTPTIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVLGGFVTAAGITLGMDELYK 239

RESULT 3  
AAG65781  
ID AAG65781 standard; protein; 893 AA.

XX AAG65781;

XX 07-JAN-2002 (first entry)

DE Amino acid sequence of HSPDE4A1-E222G fusion protein.

XX PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;  
KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;  
XX fusion protein.

OS Homo sapiens.  
OS Aequorea victoria.

XX WO200179526-A2.

XX 25-OCT-2001.

XX 11-APR-2001; 2001WO-DK000264.

XX 17-APR-2000; 2000DK-00000651.

XX 29-MAY-2000; 2000DK-00000845.

XX (BIOI-) BIOIMAGE AS.

XX

PI Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;  
PI Praestegaard M;  
XX  
XX  
DR WPI: 2001-611727/70.  
DR N-PSDB; AAI66852.  
XX  
XX  
PT Determining if a compound is a dislocator of PDE4 for identifying  
PT compounds for treating CNS and inflammatory disease comprises identifying  
PT compounds which remove PDE4 spots.  
XX  
XX  
PS Example 1; Page 156-160; 160pp; English.  
XX  
CC The invention relates to determining, if a compound, is a dislocator of  
CC PDE4. The method comprises testing if the compound removes PDE4 spots,  
CC which may optionally be induced by a Rolipram-like reference compound,  
CC and testing if it inhibits the catalytic activity of the PDE4, where the  
CC compound is a dislocator of PDE4, if it removes PDE spots and if it does  
CC not inhibit the catalytic activity of PDE4. The method is useful for  
CC identifying compounds useful for the treatment of diseases of the central  
CC nervous system such as depression and for the treatment of inflammatory  
CC disease such as joint inflammation, Crohn's disease, inflammatory bowel  
CC disease, respiratory diseases, chronic obstructive pulmonary disease  
CC (COPD), including asthma, chronic bronchitis, pulmonary emphysema,  
CC endotoxemic shock, toxic shock syndrome, systemic lupus erythematosus,  
CC psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV  
CC infection. The use of a reagent that can mimic or reverse the effect of  
CC the compound with affinity for the catalytic site on intracellular  
CC distribution of the PDE for the preparation of a medicament. The present  
CC sequence represents the amino acid sequence of a HSPDE4A1-E222G fusion  
CC protein  
XX  
SQ Sequence 893 AA;  
  
Query Match 99.2%; Score 1266; DB 4; Length 893;  
Best Local Similarity 99.8%; Pred. No. 9.6e-122;  
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 MYSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60  
Db 655 MYSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 714  
  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120  
Db 715 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 774  
  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
Db 775 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 834  
  
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239  
Db 835 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 893  
  
RESULT 4  
AAG65782  
ID AAG65782 standard; protein; 1132 AA.  
XX  
AC AAG65782;  
XX  
XX  
DT 07-JAN-2002 (first entry)  
XX  
DE Amino acid sequence of HSPDE4A4-E222G fusion protein.  
XX  
XX PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;  
KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;  
KW fusion protein.  
XX  
OS Homo sapiens.  
OS Aequorea victoria.  
XX  
XX WO200179526-A2.  
PN  
XX

PD 25-OCT-2001.  
XX  
XX  
PF 11-APR-2001; 2001WO-DK000264.  
XX  
XX  
PR 17-APR-2000; 2000DK-00000651.  
PR 29-MAY-2000; 2000DK-00000849.  
XX  
XX  
PA (BIOI-) BIOIMAGE AS.  
XX  
XX  
PI Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;  
PI Praestegaard M;  
XX  
XX  
DR WPI: 2001-611727/70.  
DR N-PSDB; AAI66853.  
XX  
XX  
PT Determining if a compound is a dislocator of PDE4 for identifying  
PT compounds for treating CNS and inflammatory disease comprises identifying  
PT compounds which remove PDE4 spots.  
XX  
XX  
PS Example 1; Page 162-167; 160pp; English.  
XX  
CC The invention relates to determining, if a compound, is a dislocator of  
CC PDE4. The method comprises testing if the compound removes PDE4 spots,  
CC which may optionally be induced by a Rolipram-like reference compound,  
CC and testing if it inhibits the catalytic activity of the PDE4, where the  
CC compound is a dislocator of PDE4, if it removes PDE spots and if it does  
CC not inhibit the catalytic activity of PDE4. The method is useful for  
CC identifying compounds useful for the treatment of diseases of the central  
CC nervous system such as depression and for the treatment of inflammatory  
CC disease such as joint inflammation, Crohn's disease, inflammatory bowel  
CC disease, respiratory diseases, chronic obstructive pulmonary disease  
CC (COPD), including asthma, chronic bronchitis, pulmonary emphysema,  
CC endotoxemic shock, toxic shock syndrome, systemic lupus erythematosus,  
CC psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV  
CC infection. The use of a reagent that can mimic or reverse the effect of  
CC the compound with affinity for the catalytic site on intracellular  
CC distribution of the PDE for the preparation of a medicament. The present  
CC sequence represents the amino acid sequence of a HSPDE4A4-E222G fusion  
CC protein  
XX  
SQ Sequence 1132 AA;  
  
Query Match 99.2%; Score 1266; DB 4; Length 1132;  
Best Local Similarity 99.6%; Pred. No. 1.4e-121;  
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 MYSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60  
Db 894 MYSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 953  
  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120  
Db 954 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 1013  
  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
Db 1014 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 1073  
  
QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239  
Db 1074 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 1132  
  
RESULT 5  
AAE17517  
ID AAE17517 standard; protein; 239 AA.  
XX  
AC AAE17517;  
XX  
XX  
DT 22-APR-2002 (first entry)  
XX  
XX Enhanced F64L jellyfish green fluorescent protein mutant.  
DE  
XX

KW	Jellyfish; green fluorescent protein; GFP; protein redistribution;	
KW	cellular function; genetic reporter; mutant; Stoke's shift; mitein.	
XX		
OS	Aequorea victoria.	
OS	Synthetic.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 65	/note= "Wild type Phe substituted with Leu; This
FT		corresponds to position 64 in the wild type protein"
FT		
XX		
PN	WO200198338-A2.	
XX		
PD	27-DEC-2001.	
XX		
XX	18-JUN-2001; 2001WO-EP006848.	
XX		
PR	19-JUN-2000; 2000DK-00000953.	
PR	20-JUN-2000; 2000US-0212681P.	
PR	10-MAY-2001; 2001DK-00000739.	
PR	10-MAY-2001; 2001US-0290170P.	
XX		
PA	(BIOI-) BIOIMAGE AS.	
XX		
PI	Bjorn SP, Pagliaro L, Thastrup O;	
XX		
DR	NP; 2002-098224/13.	
DR	W-PSDE; AAD28162.	
XX		
PT	Novel fluorescent protein in in vitro assay for measuring protein kinase	
PT	activity or dephosphorylation activity, or for measuring protein	
PT	redistribution, has a green fluorescent protein with F64L and E222G	
PT	mutation.	
XX		
PS	Example 1; Page 35; 41pp; English.	
XX		
CC	The invention relates to a fluorescent protein derived from green	
CC	fluorescent protein (GFP) or its analogue. The GFP containing mutations	
CC	at F64L and E222G has a bigger compared to other GFP's making it very	
CC	suitable for high throughput screening due to better resolution. The	
CC	fluorescent protein is useful in invitro assays for measuring protein	
CC	kinase activity or dephosphorylation activity, or for measuring protein	
CC	redistribution. The fluorescent protein is useful in studying cellular	
CC	functions in living cells; as protein tags in transgenic animals, living	
CC	and fixed cells; organelle tags, secretion marker and genetic reporter.	
CC	The fluorescent protein is also useful as a cell or organelle integrity	
CC	marker, a marker for changes in cell morphology, as transfection marker,	
CC	and as a marker to be used in combination with fluorescence activated	
CC	cell sorting (FACS). The novel proteins can also be used as reporters to	
CC	monitor live or dead biomass of organisms, such as fungi. The fluorescent	
CC	protein is also useful as markers in transcriptional and translational	
CC	fusions for performing transposon vector mutagenesis and as a reporter	
CC	useful for bacterial detection. Transposons encoding the fluorescent protein are	
CC	useful for screening promoters and for tagging plasmids and chromosomes.	
CC	The fluorescent protein engineered into the genome of a phage is useful	
CC	for designing diagnostic tool. The present sequence is enhanced F64L	
CC	jellyfish green fluorescent protein (GFP) mutant	
XX		
SQ	Sequence 239 AA;	
	Query Match	
	Best Local Similarity 98.6%; Score 1258; DB 5; Length 239;	
	Matches 237; Conservative 99.2%; Pred. No. 9.4e-122;	
	Mismatches 2; Indels 0; Gaps 0;	
Qy	1 MVSGBELFTGVWPILVELDGVNGHKFSVSGEGEGATYKGLTLKFCITCTGKLPVWPPT 60	
Db	1 MVSGBELFTGVWPILVELDGVNGHKFSVSGEGEGATYKGLTLKFCITCTGKLPVWPPT 60	
Qy	61 LVTTLSVGVCFSYRPHMKQHDFFKSNMPGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120	
Db	61 LVTTLSVGVCFSYRPHMKQHDFFKSNMPGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120	
Qy	121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180	

CC biosensors of the invention can be used to investigate a wide range of  
CC cellular activities and to screen compounds which modulate these  
CC activities. Biosensors containing a recognition site for caspase, for  
CC example, may be used for the screening of compounds which modulate  
CC apoptosis, while biosensors containing other protease recognition sites  
CC may be used for the detection of proteolytic toxins (such as anthrax  
CC lethal factor). The method provides improved target validation and  
CC candidate compound optimisation by combining many cell screening formats  
CC with fluorescence-based molecular reagents and computer-based feature  
CC extraction, data analysis and automation, resulting in increased quantity  
CC and speed of data collection and faster evaluation of drug candidates.  
CC Sequences AAB22881-B22885 represent fluorescent proteins which may be used  
CC as components of biosensor fusion proteins of the invention  
XX  
SQ Sequence 239 AA;

Query Match 98.4%; Score 1255; DB 3; Length 239;  
Best Local Similarity 98.7%; Pred. No. 1.9e-121;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MYSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60  
DB 1 MYSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60  
QY 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239

RESULT 7  
AAV54349  
ID AAY54349 standard; protein; 239 AA.  
XX AAY54349;  
XX AC AAY54349;  
XX DT 06-APR-2000 (first entry)  
DE Amino acid sequence of the mutant green fluorescent protein EGFP.  
XX Fluorescent protein; green fluorescent protein; emission intensity;  
KW fluorescence; pH detection; pH sensor; EGFP.  
XX Synthetic.  
OS Aequorea victoria.

Key Location/Qualifiers  
FH Misc-difference 65  
FT /note= "wild type Phe substituted with Leu"  
FT Misc-difference 66 /note= "wild type Ser substituted with Thr"  
FT Misc-difference 232 /note= "wild type His substituted with Leu"  
XX WO9964592-A2.  
XX 16-DEC-1999.  
XX 08-JUN-1999; 99WO-US012850.  
XX 09-JUN-1998; 98US-00094359.  
XX 13-OCT-1998; 98US-00172063.  
XX (REGC ) UNIV CALIFORNIA.  
XX (UYOR-) UNIV OREGON STATE.

PI Tsien RV, Llopis J, Wachtler RM;  
XX WPI; 2000-116540/10.  
DR N-PSDB; AA245642.  
XX New functional engineered green fluorescent proteins, used for measuring  
PT the pH in biological samples and cells.  
XX Disclosure; Page 9; 89pp; English.  
XX The present sequence represents a functional engineered fluorescent  
CC protein based on the Aequorea green fluorescent protein (GFP). The  
CC emission intensity changes as pH varies between 5 and 10 of the present  
CC protein are novel. The functional engineered fluorescent proteins show  
CC reversible changes in fluorescence over physiological pH ranges. They can  
CC be used for determining the pH of samples and cells. The polynucleotides  
CC can also be used to produce transgenic animals. The fluorescent protein  
CC pH sensors can be delivered to cells in the form of polynucleotides  
CC encoding the protein sensor fused to a targeting signal. The targeting  
CC signal directs the expression of the protein sensors to restricted cell  
CC locations. This makes it possible to measure the pH of a precisely  
CC defined cellular region or organelle  
XX  
SQ Sequence 239 AA;

Query Match 98.4%; Score 1255; DB 3; Length 239;  
Best Local Similarity 98.7%; Pred. No. 1.9e-121;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MYSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60  
DB 1 MYSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60  
QY 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239

RESULT 8  
AAV79584  
ID AAY79584 standard; peptide; 239 AA.  
XX AAY79584;  
XX AC AAY79584;  
XX DT 29-AUG-2000 (first entry)  
XX EGFP signal domain.  
DE EGFP signal domain.  
XX Protease; biosensor; EGFP; signal peptide; cell screening; assay;  
KW analysis; drug discovery.  
XX Unidentified.  
OS WO200026408-A2.  
XX 11-MAY-2000.  
XX 29-OCT-1999; 99WO-US025431.  
XX 30-OCT-1998; 98US-0106308P.  
XX 26-MAY-1999; 99US-0136078P.  
XX (CELL-) CELLOMICS INC.  
XX Guiliano KA, Bright G, Olson K, Burroughs-Tencza S;

XX	WPI; 2300-365644/31.	PF	17-MAY-2000; 2000WO-US013684.	XX
DR	N-PSDB; AAA27573.	XX		
XX	Recombinant nucleic acid encoding a protease biosensor useful for	PR	21-MAY-1993; 99US-00316919.	
PT	fluorescence based cell and molecular biochemical assays for drug	PR	21-MAY-1993; 99US-00316920.	
PT	discovery comprising three operably linked nucleic acid sequences.	XX		
XX		PA	(REGC ) UNIV CALIFORNIA.	
XX		XX		
PS	Claim 14; Fig 29A; 218pp; English.	PI	Tsien RV, Baird GA;	
XX		XX		
CC	The present sequence is that of the EGFP signal domain, which can be	DR	WPI; 2001-032017/04.	
CC	included in novel recombinant protease biosensors (PBs) of the invention.	DR	N-PSDB; AAC90488.	
CC	The PBs (see AAY79638-54) comprise: a first domain (see AAY79579-87)	XX		
CC	comprising at least 1 detectable polypeptide signal such as the present	PT	Novel fluorescent proteins comprising a sensor protein inserted into	
CC	sequence; a second domain (see AAY79588-622) comprising at least 1	PT	them, useful for measuring the response of a sensor biological, chemical,	
CC	protease recognition site; and a third domain (see AAY79623-37)	XX	electrical or physiological parameter in vivo or in vitro.	
CC	comprising at least 1 reacting target sequence. A recombinant nucleic	PS	Disclosure; Page 24; 94pp; English.	
CC	acid (see AAA27627-43) encoding the PB, an expression vector, and a	XX		
CC	genetically engineered host cell are also claimed. A claimed method for	CC	The present sequence is a fluorescent protein used in the construction of	
CC	identifying compounds that modify protease activity in a cell involves	CC	a fluorescent protein indicator. The indicator comprises a sensor	
CC	contacting a host cell that possesses the recombinant PB with a test	CC	polypeptide that is responsive to a chemical, biological, electrical or	
CC	compound, and determining the PB distribution in the host cell, where	CC	physiological parameter, and a fluorescence protein functional group. The	
CC	changes in the distribution of the PB are correlated with modification of	CC	sensor polypeptide is operatively inserted into the fluorescent moiety.	
CC	protease activity by the test compound. Claimed kits for identifying	CC	The fluorescent indicator is useful for detecting the presence of a	
CC	compounds that modify protease activity in a host cell include the	CC	response inducing member in a sample. The method involves contacting the	
CC	recombinant nucleic acid, or the recombinant PB, or the vector, or the	CC	sample with the indicator and detecting a change in fluorescence, in	
CC	host cell. The PB is useful in high content screens to detect in vivo	CC	which a change is indicative of the effect of the parameter on the sensor	
CC	activation of enzymatic activity, and to identify specific activity based	CC	polypeptide. The novel fluorescent proteins are advantageous due to their	
CC	on cleavage of a known recognition motif	CC	reduced size as compared to the FRET (fluorescence resonance energy	
XX		CC	transfer)-based sensors	
XX		XX		
SQ	Sequence 239 AA;	SQ	Sequence 239 AA;	
<hr/>				
Query Match 98.4%; Score 1255; DB 3; Length 239;				
Best Local Similarity 98.7%; Pred. No. 1.9e-121;				
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;				
QY	1 MYSKGELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGKLTFLKFICTTGKLPVPWPT 60	QY	1 MYSKGELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGKLTFLKFICTTGKLPVPWPT 60	
Db	1 MYSKGELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGKLTFLKFICTTGKLPVPWPT 60	Db	1 MYSKGELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGKLTFLKFICTTGKLPVPWPT 60	
QY	61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120	QY	61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120	
Db	61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120	Db	61 LVTTLTSGVQCFSRYPDHMKQHDFFKSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120	
QY	121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180	QY	121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180	
Db	121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180	Db	121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIMADKQNGIKVNFKIRHNIEDGSVQLA 180	
QY	181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239	QY	181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239	
Db	181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239	Db	181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGFTVTAAGITLGMDELYK 239	
<hr/>				
RESULT 9				
AAB50804				
ID	AAB50804 standard; protein; 239 AA.	ID	AAB85900 standard; protein; 239 AA.	
XX		XX		
AC	AAB50804;	AC	AAB85900;	
XX		XX		
DT	14-MAR-2001 (first entry)	DT	30-NOV-2001 (first entry)	
XX		XX		
DE	Jellyfish GFP mutant EGFP.	DE	A. victoria green fluorescent protein (GFP) and linker sequence.	
XX		XX		
KW	Aequorea victoria; jellyfish; fluorescent protein indicator;	XX		
KW	green fluorescent protein; GFP; linker moiety; sensor;	KW	Melanin concentrating hormone receptor; MCHR; MCH; chimeric; fusion;	
KW	calmodulin-binding domain; mutant; mutein.	KW	fluorescent polypeptide; orexigenic; anabolic; food intake; GFP;	
XX		KW	green fluorescent protein.	
OS	Aequorea victoria.	OS	Synthetic.	
XX		OS	Aequorea victoria.	
PN	WO200071565-A2.	PN	WO200168706-A1.	
XX		XX		
PD	30-NOV-2000.	XX		
XX		XX		



PD 20-SEP-2001.  
 XX  
 PF 14-MAR-2001; 2001WO-US008071.  
 XX  
 PR 15-MAR-2000; 2000US-0189698P.  
 XX  
 PA (MERI ) MERCK & CO INC.  
 XX  
 PI Marsh DJ;  
 XX  
 PI WPI; 2001-565791/63.  
 DR N-PSDB; AAH47304.  
 DR  
 XX Fusion proteins comprising melanin concentrating hormone receptor  
 PT peptides and fluorescent proteins, useful for identifying appetite  
 PT stimulants.  
 XX  
 PS Claim 2; Page 14; 71pp; English.  
 CC The invention provides melanin concentrating hormone (MCH) receptor  
 CC (MCHR) chimeric and fusion proteins. The MCHR chimeric proteins comprise  
 CC MCHR polypeptide regions from different species. The MCHR fusion protein  
 CC comprise MCHR polypeptide region and a fluorescent polypeptide region  
 CC joined directly, or via a linker, to the carboxy side of the MCHR  
 CC polypeptide region. The MCHR fusion proteins can be expressed by standard  
 CC recombinant methodology. MCH action promotes feeding (orexigenic) and up  
 CC regulation of MCH activity stimulates food intake. The present sequence  
 CC represents a A. victoria green fluorescent protein (GFP) and a linker  
 CC sequence  
 XX  
 SQ Sequence 239 AA;  
 Query Match 98.4%; Score 1255; DB 4; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 1.9e-121;  
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MYSKGEEFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60  
 DB 1 MYSKGEEFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60  
 QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTTLTGYVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 QY 181 DHYQQTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239  
 DB 181 DHYQQTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239  
 RESULT 11  
 AAB31171  
 ID AAB31171 standard; protein; 239 AA.  
 XX  
 AC AAB31171;  
 XX  
 DT 02-APR-2001 (first entry)  
 XX  
 DE Amino acid sequence of a green fluorescent protein (GFP).  
 XX  
 KW Growth rate; death rate; reporter gene; luminescent protein;  
 KW fluorescent product; luciferase; green fluorescent protein; GFP.  
 XX  
 OS Aequorea victoria.  
 XX  
 PN WO200075367-A1.  
 XX  
 PD 14-DEC-2000.  
 XX  
 PF 07-JUN-2000; 2000WO-FI000507.

XX 07-JUN-1999; 99FI-00001296.  
 XX  
 PA (LILLI) LILIUS E.  
 PA (VIRT) VIRT M.  
 XX  
 PI Lilius E, Virta M;  
 XX  
 PI WPI; 2001-061737/07.  
 DR N-PSDB; AAC86954.  
 DR  
 XX Assessing growth and death rates of a micro-organism in a desired  
 PT environment, by introducing 2 reporter genes encoding luminescent and  
 PT fluorescent products and detecting luminescent fluorescence.  
 XX  
 PS Disclosure; Page 27; 32pp; English.  
 CC  
 CC The specification describes a method for assessing the growth rate and  
 CC death rate of a micro-organism within a predetermined time period in a  
 CC desired environment. The method comprises introducing at least two  
 CC reporter genes encoding luminescent and/or fluorescent products into the  
 CC micro-organisms, incubating the micro-organism within the desired a  
 CC environment, and detecting luminescence and/or fluorescence after a  
 CC predetermined time period. Use of two different markers within a micro-  
 CC organism enables the differentiation between growth and death rates. The  
 CC method is used to assess the growth rate and death rate of a micro-  
 CC organism within a predetermined time period in a desired environment. The  
 CC present sequence represents a green fluorescent protein (GFP), and is  
 CC encoded by a plasmid which encodes luminescent and fluorescent proteins,  
 CC and is used in the method of the invention  
 XX  
 SQ Sequence 239 AA;  
 Query Match 98.4%; Score 1255; DB 4; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 1.9e-121;  
 Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MYSKGEEFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60  
 DB 1 MYSKGEEFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60  
 QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTTLTGYVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 QY 181 DHYQQTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239  
 DB 181 DHYQQTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239  
 RESULT 12  
 AAG66198  
 ID AAG66198 standard; protein; 239 AA.  
 XX  
 AC AAG66198;  
 XX  
 DT 17-JUN-2002 (first entry)  
 XX  
 DE A. victoria green fluorescent protein (EGFP).  
 XX  
 KW Cyan-green fluorescent protein; fluorescence; recombinant; GFP;  
 KW green fluorescent protein; EGFP.  
 XX  
 OS Aequorea victoria.  
 XX  
 PN JP2002045189-A.  
 XX  
 PD 12-FEB-2002.  
 XX



XX FH Key Location/Qualifiers  
FT Misc-difference 1. .3 /note= "Wild-type GFP Met-Ser are replaced with Met-Val-Ser"  
FT FT  
FT FT  
FT Misc-difference 65 /note= "GFP Phe64 is replaced by Leu"  
FT FT  
FT Misc-difference 66 /note= "GFP Ser65 is replaced by Thr"  
XX XX  
PN EP1178109-A1.  
XX 06-FEB-2002.  
XX 03-AUG-2001; 2001EP-00306650.  
PR 04-AUG-2000; 2000JP-00237166.  
XX XX  
PA (RIKE ) RIKEN KK.  
XX Miyawaki A, Sawano A;  
PI WPI: 2002-208112/27.  
XX N-PSDB; AAD27910.  
XX  
PT Method for mutagenesis, e.g. for introducing certain or random mutations at certain sites of the nucleotide sequence, comprises synthesizing a mutated strand and a complementary strand by use of megaprimer.  
PT  
PS Example 1; Page 13-14; 31pp; English.  
XX  
XX The invention relates to a method for mutagenesis that comprises synthesizing a mutated strand and a complementary strand by use of megaprimer. The method basically comprises a DNA synthesis in which one or more primers that have a nucleotide sequence containing at least one mutation and a phosphorylated 5'-terminus are annealed to a template DNA and then subjected to an elongation reaction using a thermostable high-fidelity DNA polymerase, after which the phosphorylated 5'-terminus and the elongated terminus are ligated by means of a thermostable DNA ligase to synthesise a circular DNA containing the primers; a digestion in which the step of DNA synthesis is repeated several times to amplify the DNA containing the primers and then, at least DNAs other than the amplified circular DNA are digested into several fragments; and a double-stranded DNA synthesis in which, with the several fragments obtained in the step of digestion as megaprimer, the megaprimer are annealed to the circular DNA synthesised above, followed by an elongation reaction performed using the thermostable high-fidelity DNA polymerase. The method is useful for mutagenesis, particularly for introducing certain mutations at certain sites of a nucleotide sequence, or for introducing random mutations at certain sites of the nucleotide sequence. The present method is simple, speedy, economical and widely applicable. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used for mutagenesis in an exemplification of the invention. The EGFP is derived by altering the green fluorescent protein (GFP) sequence of Aequorea victoria  
XX  
SQ Sequence 239 AA;

Query Match  
Best Local Similarity 98.4%; Score 1255; DB 5; Length 239;  
Matches 236; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY 1 MYSKGEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTCLKFICTTGKLPVPWPT 60  
DB 1 MYSKGEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTCLKFICTTGKLPVPWPT 60  
QY 61 LVTLLSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTLLTYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNPKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNPKIRHNIEDGSVQLA 180

QY 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGGFVTAAGITLGMDELYK 239  
DB 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGGFVTAAGITLGMDELYK 239

RESULT 15  
AAE34958  
ID AAE34958 standard; protein; 239 AA.  
XX AAE34958;  
XX 28-MAY-2003 (first entry)  
XX Aequorea victoria enhanced green fluorescent protein (EGFP).  
XX Phosphorylation indicator; fluorescent protein; detection; phosphatase;  
XX Kinase; enhanced green fluorescent protein; EGFP.  
XX Aequorea victoria.  
XX WO200295058-A2.  
XX 28-NOV-2002.  
XX 24-MAY-2002; 2002WO-US016955.  
XX 24-MAY-2001; 2001US-00865291.  
XX (REGC ) UNIV CALIFORNIA.  
XX Tsien RY, Ting AY, Zhang J;  
XX WPI: 2003-148474/14.  
XX N-PSDB; AAD53428.  
XX Novel chimeric phosphorylation indicators, useful for detecting kinase/phosphatase in samples, has donor molecule, phosphorylatable domain, phosphoaminoacid binding domain, and acceptor molecule, in operative linkage.  
XX Disclosure; Col 56-57; 38pp; English.  
XX The present invention relates to chimeric phosphorylation indicators comprising a phosphorylation polypeptide and a fluorescent protein or in operative linkage, a donor molecule, a phosphorylatable domain, a phosphoaminoacid binding domain (PAABD) and an acceptor molecule. The phosphorylation indicators of the invention are useful for detecting kinases or phosphatases in a biological sample. They are also useful in high throughput analysis e.g. for detecting a kinase inhibitor or phosphatase inhibitor. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used in the invention  
XX  
SQ Sequence 239 AA;

Query Match  
Best Local Similarity 98.4%; Score 1255; DB 6; Length 239;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MYSKGEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTCLKFICTTGKLPVPWPT 60  
DB 1 MYSKGEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTCLKFICTTGKLPVPWPT 60  
QY 61 LVTLLSYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTLLTYGQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNPKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNPKIRHNIEDGSVQLA 180  
QY 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGGFVTAAGITLGMDELYK 239  
DB 181 DHYQQNTPIGDGFPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGGFVTAAGITLGMDELYK 239

Db 131 DHYQNTPIGDFVLLPDNHYLSTQSALSKDPNEKRDHMLLEFVTAGITLGMDELYK 239

Search completed: June 21, 2004, 15:54:13  
Job time : 47.1111 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:55 ; Search time 12.7778 Seconds  
(without alignments)  
965.630 Million cell updates/sec

Title: US-09-887-784-222G

Perfect score: 1276

Sequence: 1 MVSKGELFTGVVPIVLVD.....VLGGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA.\*

- 1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/prodata/2/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/prodata/2/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/prodata/2/iaa/6C\_COMB.pep.\*
- 6: /cgn2\_6/prodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1255	98.4	239	3	US-09-172-063-3
2	1255	98.4	239	4	US-09-513-783A-46
3	1255	98.4	239	4	US-09-316-919-4
4	1255	98.4	239	4	US-09-602-641-3
5	1255	98.4	239	4	US-09-920-922-2
6	1255	98.4	281	3	US-09-062-102-1
7	1255	98.4	281	4	US-09-364-946-1
8	1255	98.4	294	4	US-09-513-783A-2
9	1255	98.4	323	3	US-09-172-063-21
10	1255	98.4	323	4	US-09-602-641-21
11	1255	98.4	364	3	US-09-085-305-6
12	1255	98.4	379	4	US-09-417-197-129
13	1255	98.4	434	4	US-09-800-170-48
14	1255	98.4	442	4	US-09-417-197-127
15	1255	98.4	459	4	US-09-513-783A-170
16	1255	98.4	544	4	US-09-417-197-113
17	1255	98.4	544	4	US-09-417-197-115
18	1255	98.4	604	4	US-09-417-197-59
19	1255	98.4	605	4	US-09-417-197-41
20	1255	98.4	606	4	US-09-417-197-65
21	1255	98.4	607	4	US-09-417-197-47
22	1255	98.4	630	4	US-09-417-197-63
23	1255	98.4	631	4	US-09-417-197-39
24	1255	98.4	633	4	US-09-417-197-45
25	1255	98.4	635	4	US-09-417-197-125
26	1255	98.4	642	2	US-08-818-253-2
27	1255	98.4	642	2	US-08-818-253-6

28	1255	98.4	642	3	US-08-818-252-2	Sequence 2, Appli
29	1255	98.4	642	3	US-08-818-252-6	Sequence 6, Appli
30	1255	98.4	652	2	US-08-818-253-4	Sequence 4, Appli
31	1255	98.4	652	3	US-08-818-252-4	Sequence 4, Appli
32	1255	98.4	718	4	US-09-417-197-75	Sequence 75, Appli
33	1255	98.4	719	4	US-09-417-197-51	Sequence 51, Appli
34	1255	98.4	726	4	US-09-417-197-71	Sequence 71, Appli
35	1255	98.4	727	4	US-09-417-197-139	Sequence 139, App
36	1255	98.4	783	4	US-09-513-783A-176	Sequence 176, App
37	1255	98.4	797	4	US-09-417-197-141	Sequence 141, App
38	1255	98.4	797	4	US-09-417-197-143	Sequence 143, App
39	1255	98.4	798	4	US-09-417-197-77	Sequence 77, Appli
40	1255	98.4	805	4	US-09-513-783A-178	Sequence 178, App
41	1255	98.4	806	4	US-09-417-197-53	Sequence 53, Appli
42	1255	98.4	836	4	US-09-417-197-61	Sequence 61, Appli
43	1255	98.4	842	4	US-09-417-197-43	Sequence 43, Appli
44	1255	98.4	843	4	US-09-417-197-117	Sequence 117, App
45	1255	98.4	853	4	US-09-417-197-119	Sequence 119, App

ALIGNMENTS

RESULT 1  
US-09-172-063-3  
; Sequence 3, Application US/09172063  
; Patent No. 6150176  
; GENERAL INFORMATION:  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Miyawaki, Atsushi  
; APPLICANT: Lloptis, Juan  
; APPLICANT: Wachter, Rebekka M.  
; APPLICANT: Remington, S. James  
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
; FILE REFERENCE: 07257/071001  
; CURRENT APPLICATION NUMBER: US/09/172,063  
; CURRENT FILING DATE: 1998-10-13  
; EARLIER APPLICATION NUMBER: 09/094,359  
; EARLIER FILING DATE: 1998-06-09  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: EGFP  
US-09-172-063-3

Query Match	98.4%;	Score	1255;	DB 3;	Length	239;			
Best Local Similarity	98.7%;	Pred. NO.	2.5e-127;						
Matches	236;	Conservative	1;	Mismatches	2;	Indels	0;	Gaps	0;
Qy	1	MVSKGELFTGVVPIVLVDGDVNGHKFSVSGEGDATYVKLTLLKFICTTGKLPVWPWT	60						
Db	1	MVSKGELFTGVVPIVLVDGDVNGHKFSVSGEGDATYVKLTLLKFICTTGKLPVWPWT	60						
Qy	61	LVTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNKYKTRAEVKFEGDTL	120						
Db	61	LVTLLTYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNKYKTRAEVKFEGDTL	120						
Qy	121	VNRIELKGDIFDEKGNILGHKLEYNYSNHNVIYIMADKQNGIKVNFIRHNIEDGSVQLA	180						
Db	121	VNRIELKGDIFDEKGNILGHKLEYNYSNHNVIYIMADKQNGIKVNFIRHNIEDGSVQLA	180						
Qy	181	DHYQONTPTIGDGPVLLPDNHHYLSSTQSALSCKDPNEKRDHNVLGGFVTAAGITLGMDELYK	239						
Db	181	DHYQONTPTIGDGPVLLPDNHHYLSSTQSALSCKDPNEKRDHNVLGGFVTAAGITLGMDELYK	239						

```
RESULT 2
US-09-513-783A-46
; Sequence 46, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 46
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGFP
US-09-513-783A-46

Query Match      98.4%; Score 1255; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.5e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFVTAAGITLGMDELYK 239

RESULT 3
US-09-316-919-4
; Sequence 4, Application US/09316919
; Patent No. 6469154
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: 07257/073001
; CURRENT APPLICATION NUMBER: US/09/316,919
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-316-919-4

Query Match      98.4%; Score 1255; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.5e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFVTAAGITLGMDELYK 239

RESULT 4
US-09-602-641-3
; Sequence 3, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Ilopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-602-641-3

Query Match      98.4%; Score 1255; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.5e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGFVTAAGITLGMDELYK 239

RESULT 5
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. 6673610
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match      98.4%; Score 1255; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.5e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGKLTALKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGKLTALKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQDPNEKRDHVMVLFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQDPNEKRDHVMVLFVTAAGITLGMDELYK 239

RESULT 6
US-09-062-102-1
; Sequence 1, Application US/09062102
; Patent No. 6130313
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangqiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100
; CURRENT APPLICATION NUMBER: US/09/062,102
; CURRENT FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: US 60/060,855
; EARLIER FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
US-09-062-102-1

Query Match      98.4%; Score 1255; DB 3; Length 281;
Best Local Similarity 98.7%; Pred. No. 3.1e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGKLTALKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGKLTALKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQDPNEKRDHVMVLFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQDPNEKRDHVMVLFVTAAGITLGMDELYK 239

RESULT 7
US-09-364-946-1
; Sequence 1, Application US/09364946
; Patent No. 6306600
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangqiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100CJP/D2
; CURRENT APPLICATION NUMBER: US/09/364,946
; CURRENT FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: US 09/191,233
; EARLIER FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
US-09-364-946-1

Query Match      98.4%; Score 1255; DB 4; Length 281;
Best Local Similarity 98.7%; Pred. No. 3.1e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGKLTALKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGKLTALKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLTGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQDPNEKRDHVMVLFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQDPNEKRDHVMVLFVTAAGITLGMDELYK 239

RESULT 8
US-09-513-783A-2
; Sequence 2, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GFP-DEVD-Annexin II construct
US-09-513-783A-2

Query Match      98.4%; Score 1255; DB 4; Length 294;
Best Local Similarity 98.7%; Pred. No. 3.3e-127;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGKLTALKFICTTGKLPVPWPT 60
|||||
```

Db 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTAKFICTTGKLPVWPWT 60  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRLEKGDIFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
Db 121 VNRLEKGDIFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239

## RESULT 9

US-09-172-063-21

; Sequence 21, Application US/09172063

; Patent No. 6150176

; GENERAL INFORMATION:

; APPLICANT: Tsiens, Roger Y.

; APPLICANT: Miyawaki, Atsushi

; APPLICANT: Llopi, Juan

; APPLICANT: Wachter, Rebekka M.

; APPLICANT: Remington, S. James

; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR

; FILE REFERENCE: MEASURING THE PH OF A BIOLOGICAL SAMPLE

; CURRENT APPLICATION NUMBER: US/09/172,063

; CURRENT FILING DATE: 1998-10-13

; EARLIER APPLICATION NUMBER: 09/094,359

; EARLIER FILING DATE: 1998-06-09

; NUMBER OF SEQ ID NOS: 38

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 21

; LENGTH: 323

; TYPE: PRT

; ORGANISM: Aequorea victoria

; FEATURE:

; NAME/KEY: VARIANT

; LOCATION: (0)...(0)

; OTHER INFORMATION: GT-EGFP

US-09-172-063-21

Query Match 98.4%; Score 1255; DB 3; Length 323;

Best Local Similarity 98.7%; Pred. No. 3.8e-127;

Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTAKFICTTGKLPVWPWT 60  
Db 85 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTAKFICTTGKLPVWPWT 144  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 145 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 204  
QY 121 VNRLEKGDIFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
Db 205 VNRLEKGDIFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 264  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239  
Db 265 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGGFVTAAGITLGMDELYK 323

## RESULT 10

US-09-602-641-21

; Sequence 21, Application US/09602641

; Patent No. 6608189

; GENERAL INFORMATION:

; APPLICANT: Tsiens, Roger Y.

; APPLICANT: Miyawaki, Atsushi

; APPLICANT: Llopi, Juan

; APPLICANT: Wachter, Rebekka M.  
; APPLICANT: Remington, S. James  
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
; FILE REFERENCE: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
; CURRENT APPLICATION NUMBER: US/09/602,641  
; CURRENT FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 09/172,063  
; PRIOR FILING DATE: 1998-10-13  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 21  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
; FEATURE:

; NAME/KEY: VARIANT

; LOCATION: (0)...(0)

; OTHER INFORMATION: GT-EGFP

US-09-602-641-21

Query Match 98.4%; Score 1255; DB 4; Length 323;

Best Local Similarity 98.7%; Pred. No. 3.8e-127;

Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTAKFICTTGKLPVWPWT 60  
Db 85 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYVKLTAKFICTTGKLPVWPWT 144  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 145 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 204  
QY 121 VNRLEKGDIFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
Db 205 VNRLEKGDIFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 264  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGGFVTAAGITLGMDELYK 239  
Db 265 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGGFVTAAGITLGMDELYK 323

## RESULT 11

US-09-085-305-6

; Sequence 6, Application US/09085305

; Patent No. 6191269

; GENERAL INFORMATION:

; APPLICANT: Pollock, Allan

; APPLICANT: Lovett, David H.

; APPLICANT: Turck, Johanna

; TITLE OF INVENTION: Selective Induction of Apoptosis in

; TITLE OF INVENTION: Malignant Cancer Cells by Delivery of N-Terminal

; TITLE OF INVENTION: Interleukin-1-Alpha Pro-Piece Polypeptide

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Bozicevic &amp; Reed, LLP

; STREET: 285 Hamilton Ave, Suite 200

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94301

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/085,305

; FILING DATE: 29-MAY-1998

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:



ATTORNEY/AGENT INFORMATION:  
NAME: Francis, Carol L  
REGISTRATION NUMBER: 36,513  
REFERENCE/DOCKET NUMBER: 6510/102US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-327-3400  
TELEFAX: 650-327-3231  
TELEX:  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 364 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-085-305-6

Query Match 98.4%; Score 1255; DB 3; Length 364;  
Best Local Similarity 98.7%; Pred. No. 4.6e-127;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 60  
DB 126 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 185

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 186 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 245

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVMADKQNGIKVNFKIRHNIEDGSGVQLA 180  
DB 246 VNRIELKGIDFKEDGNILGHKLEYNHNVMADKQNGIKVNFKIRHNIEDGSGVQLA 305

QY 181 DHYQONTPIGDGVPLLPDNNHLSQTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239  
DB 306 DHYQONTPIGDGVPLLPDNNHLSQTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 364

RESULT 12  
US-09-417-197-129  
Sequence 129, Application US/09417197  
Patent No. 6518021  
GENERAL INFORMATION:  
APPLICANT: Ole THASTRUP, et al.  
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An  
FILE REFERENCE: 3759-0110P  
CURRENT APPLICATION NUMBER: US/09/417,197  
CURRENT FILING DATE: 1999-10-07  
NUMBER OF SEQ ID NOS: 143  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 129  
LENGTH: 379  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: actin-binding-domain-EGFP fusion  
US-09-417-197-129

Query Match 98.4%; Score 1255; DB 4; Length 379;  
Best Local Similarity 98.7%; Pred. No. 4.9e-127;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 60  
DB 141 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 200

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 201 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 260

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVMADKQNGIKVNFKIRHNIEDGSGVQLA 180  
DB 261 VNRIELKGIDFKEDGNILGHKLEYNHNVMADKQNGIKVNFKIRHNIEDGSGVQLA 320

QY 181 DHYQONTPIGDGVPLLPDNNHLSQTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239  
DB 321 DHYQONTPIGDGVPLLPDNNHLSQTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 379

RESULT 13  
US-09-800-170-48  
Sequence 48, Application US/09800170  
Patent No. 6481667  
GENERAL INFORMATION:  
APPLICANT: Kinsella, Todd  
TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES  
FILE REFERENCE: A-68614-1/DJB/RMS/RMK  
CURRENT APPLICATION NUMBER: US/09/800,170  
CURRENT FILING DATE: 2001-08-28  
PRIOR APPLICATION NUMBER: US 60/187,130  
PRIOR FILING DATE: 2000-03-06  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 48  
LENGTH: 434  
TYPE: PRT  
ORGANISM: Synechocystis PCC6803  
US-09-800-170-48

Query Match 98.4%; Score 1255; DB 4; Length 434;  
Best Local Similarity 98.7%; Pred. No. 6e-127; 2; Indels 0; Gaps 0;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 60  
DB 196 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 255

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 256 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 315

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVMADKQNGIKVNFKIRHNIEDGSGVQLA 180  
DB 316 VNRIELKGIDFKEDGNILGHKLEYNHNVMADKQNGIKVNFKIRHNIEDGSGVQLA 375

QY 181 DHYQONTPIGDGVPLLPDNNHLSQTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239  
DB 376 DHYQONTPIGDGVPLLPDNNHLSQTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 434

RESULT 14  
US-09-417-197-127  
Sequence 127, Application US/09417197  
Patent No. 6518021  
GENERAL INFORMATION:  
APPLICANT: Ole THASTRUP, et al.  
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An  
FILE REFERENCE: 3759-0110P  
CURRENT APPLICATION NUMBER: US/09/417,197  
CURRENT FILING DATE: 1999-10-07  
NUMBER OF SEQ ID NOS: 143  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 127  
LENGTH: 442  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: EGFP-RhoA fusion  
US-09-417-197-127

Query Match 98.4%; Score 1255; DB 4; Length 442;  
Best Local Similarity 98.7%; Pred. No. 6.1e-127;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPT 60

Db 1 MVSKEELFTGVPIILVELDGVNGHKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60  
QY 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 61 LVTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
QY 181 DHYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHVLGGFVTAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHVLGGFVTAAGITLGMDELYK 239

RESULT 15  
US-09-513-783A-170  
; Sequence 170, Application US/09513783A  
; Patent No. 6416959  
; GENERAL INFORMATION:  
; APPLICANT: Giuliano, Kenneth A.  
; TITLE OF INVENTION: A System for Cell Based Screening  
; FILE REFERENCE: 97-022-11  
; CURRENT APPLICATION NUMBER: US/09/513,783A  
; CURRENT FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 170  
; LENGTH: 459  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPP-HSP27  
US-09-513-783A-170

Query Match 98.4%; Score 1255; DB 4; Length 459;  
Best Local Similarity 98.7%; Pred. No. 6.5e-127;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVPIILVELDGVNGHKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60  
Db 1 MVSKEELFTGVPIILVELDGVNGHKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60  
QY 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 61 LVTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
QY 181 DHYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHVLGGFVTAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHVLGGFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:04:05  
Job time : 12.7778 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:54:20 ; Search time 34.7778 Seconds  
(without alignments)  
1940.117 Million cell updates/sec

Title: US-09-887-784-222G  
Perfect score: 1276  
Sequence: 1 MYSKGEELFTGVVPIVLVELD.....VLGGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/PCT\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_PUB\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US05\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result; being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match %	Length	ID	Description	
1	1266	99.2	239	9	US-09-887-784-4	Sequence 4, Appli
2	1266	99.2	239	12	US-10-296-953-4	Sequence 4, Appli
3	1266	99.2	363	14	US-10-270-223-6	Sequence 6, Appli
4	1266	99.2	993	14	US-10-257-909A-30	Sequence 30, Appli
5	1266	99.2	1132	14	US-10-257-909A-32	Sequence 32, Appli
6	1258	98.6	239	9	US-09-887-784-2	Sequence 2, Appli
7	1258	98.6	239	12	US-10-296-953-2	Sequence 2, Appli
8	1258	98.4	239	9	US-09-920-922-2	Sequence 2, Appli
9	1255	98.4	239	9	US-09-999-745-4	Sequence 4, Appli
10	1255	98.4	239	10	US-09-866-538-4	Sequence 4, Appli
11	1255	98.4	239	10	US-09-797-496B-2	Sequence 2, Appli
12	1255	98.4	239	10	US-09-794-308-4	Sequence 4, Appli
13	1255	98.4	239	10	US-09-865-291-4	Sequence 4, Appli
14	1255	98.4	239	12	US-10-457-982-3	Sequence 3, Appli
15	1255	98.4	239	14	US-10-121-258-13	Sequence 13, Appli

16	1255	98.4	239	14	US-10-221-461-7	Sequence 7, Appli
17	1255	98.4	239	14	US-10-100-957A-46	Sequence 46, Appli
18	1255	98.4	239	14	US-10-177-390-2	Sequence 2, Appli
19	1255	98.4	239	14	US-10-338-411-3	Sequence 3, Appli
20	1255	98.4	239	15	US-10-370-570-4	Sequence 4, Appli
21	1255	98.4	239	15	US-10-389-640-3	Sequence 3, Appli
22	1255	98.4	259	14	US-10-314-861-11	Sequence 11, Appli
23	1255	98.4	281	12	US-09-931-232-1	Sequence 1, Appli
24	1255	98.4	288	14	US-10-314-861-37	Sequence 37, Appli
25	1255	98.4	293	14	US-10-314-861-35	Sequence 35, Appli
26	1255	98.4	294	14	US-10-100-957A-2	Sequence 2, Appli
27	1255	98.4	295	14	US-10-314-861-39	Sequence 39, Appli
28	1255	98.4	299	14	US-10-314-861-33	Sequence 33, Appli
29	1255	98.4	305	14	US-10-314-861-31	Sequence 31, Appli
30	1255	98.4	308	14	US-10-033-717-35	Sequence 35, Appli
31	1255	98.4	311	14	US-10-314-861-29	Sequence 29, Appli
32	1255	98.4	320	14	US-10-338-411-11	Sequence 11, Appli
33	1255	98.4	320	15	US-10-389-640-11	Sequence 11, Appli
34	1255	98.4	323	12	US-10-457-982-21	Sequence 21, Appli
35	1255	98.4	323	14	US-10-338-411-7	Sequence 7, Appli
36	1255	98.4	323	14	US-10-338-411-13	Sequence 13, Appli
37	1255	98.4	323	15	US-10-389-640-7	Sequence 7, Appli
38	1255	98.4	323	15	US-10-389-640-13	Sequence 13, Appli
39	1255	98.4	324	14	US-10-314-861-16	Sequence 16, Appli
40	1255	98.4	345	14	US-10-338-411-5	Sequence 5, Appli
41	1255	98.4	345	15	US-10-389-640-5	Sequence 5, Appli
42	1255	98.4	346	14	US-10-338-411-9	Sequence 9, Appli
43	1255	98.4	346	15	US-10-389-640-9	Sequence 9, Appli
44	1255	98.4	359	14	US-10-033-717-33	Sequence 33, Appli
45	1255	98.4	359	14	US-10-033-717-34	Sequence 34, Appli

ALIGNMENTS

RESULT 1  
US-09-887-784-4  
; Sequence 4, Application US/09887784  
; Patent No. US20020177189A1  
; GENERAL INFORMATION:  
; APPLICANT: BJORN, Sara et al  
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS  
; FILE REFERENCE: 3759-0115P  
; CURRENT APPLICATION NUMBER: US/09/887,784  
; CURRENT FILING DATE: 2001-06-19  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 4  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequoria Victoria  
US-09-887-784-4

Query Match		99.2%	Score 1266;	DB 9;	Length 239;
Best Local Similarity		99.6%	Pred. No. 1 se-123;		
Matches 238;		Conservative	0;	Mismatches	1;
				Indels	0;
				Gaps	0;
QY	1	MYSKGEELFTGVVPIVLVELDGVNGHKFSVSGEGDATYGKLT	FKFICTTTGKLPVPWPT	60	
DB	1	MYSKGEELFTGVVPIVLVELDGVNGHKFSVSGEGDATYGKLT	FKFICTTTGKLPVPWPT	60	
QY	61	LVTTLSYGVCESRYPDHMKQDFPKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGD	TL 120		
DB	61	LVTTLSYGVCESRYPDHMKQDFPKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGD	TL 120		
QY	121	VNRIELKGIDFDKEDGNILGHKLEYVNSHNVYIMADKQNGIKVNFKIRHNIEDGSV	QLA 180		
DB	121	VNRIELKGIDFDKEDGNILGHKLEYVNSHNVYIMADKQNGIKVNFKIRHNIEDGSV	QLA 180		
QY	181	DHYQQNTPIGDGPFVLLPDNHYLSTQSALSKDPNEKRDMVLGGFVTAAGITLGMDE	LYK 239		
DB	181	DHYQQNTPIGDGPFVLLPDNHYLSTQSALSKDPNEKRDMVLGGFVTAAGITLGMDE	LYK 239		

RESULT 2  
US-10-296-953-4  
; Sequence 4, Application US/10296953  
; Publication No. US20040072995A1  
; GENERAL INFORMATION:  
; APPLICANT: BJORN, SARA P.  
; APPLICANT: FAGLIARO, LEN  
; APPLICANT: THASTRUP, OLE  
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS  
; FILE REFERENCE: PLO095  
; CURRENT APPLICATION NUMBER: US/10/296,953  
; CURRENT FILING DATE: 2002-11-26  
; PRIOR APPLICATION NUMBER: PA 2000 00953  
; PRIOR FILING DATE: 2000-06-19  
; PRIOR APPLICATION NUMBER: 60/212,681  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: 60/290,170  
; PRIOR FILING DATE: 2001-05-10  
; PRIOR APPLICATION NUMBER: PA 2001 00739  
; PRIOR FILING DATE: 2001-05-10  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
US-10-296-953-4

Query Match 99.2%; Score 1266; DB 12; Length 239;  
Best Local Similarity 99.6%; Pred. No. 1.5e-123;  
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVGGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVGGFVTAAGITLGMDELYK 239

RESULT 3  
US-10-270-223-6  
; Sequence 6, Application US/10270223  
; Publication No. US20030143634A1  
; GENERAL INFORMATION:  
; APPLICANT: Bioclimage A/S  
; TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR COMPONENTS  
; TITLE OF INVENTION: INTERACTING CELLS, AND TO EXTRACT QUANTITATIVE INFORMATION RE  
; FILE REFERENCE: 3759-0126P  
; CURRENT APPLICATION NUMBER: US/10/270,223  
; CURRENT FILING DATE: 2002-10-11  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6  
; LENGTH: 363  
; TYPE: PRT  
; ORGANISM: Aequorea Victoria and Human  
US-10-270-223-6

Query Match 99.2%; Score 1266; DB 14; Length 363;  
Best Local Similarity 99.6%; Pred. No. 2.7e-123;  
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVGGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVGGFVTAAGITLGMDELYK 239

RESULT 4  
US-10-257-909A-30  
; Sequence 30, Application US/10257909A  
; Publication No. US20030187056A1  
; GENERAL INFORMATION:  
; APPLICANT: Bernard R. TERRY et al.  
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular  
; FILE REFERENCE: 3759-0125P  
; CURRENT APPLICATION NUMBER: US/10/257,909A  
; CURRENT FILING DATE: 2002-10-17  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 30  
; LENGTH: 893  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Fusion between Aequorea victoria and human  
US-10-257-909A-30

Query Match 99.2%; Score 1266; DB 14; Length 893;  
Best Local Similarity 99.6%; Pred. No. 9.8e-123;  
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60  
DB 655 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 714  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 715 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 774  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180  
DB 775 VNRIELKGIDFKEDGNILGHKLEYNYNHVMADKQNGIKVNFKIRHNIEDGVSQLA 834  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVGGFVTAAGITLGMDELYK 239  
DB 835 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVGGFVTAAGITLGMDELYK 893

RESULT 5  
US-10-257-909A-32  
; Sequence 32, Application US/10257909A  
; Publication No. US20030187056A1  
; GENERAL INFORMATION:  
; APPLICANT: Bernard R. TERRY et al.  
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular  
; FILE REFERENCE: 3759-0125P  
; CURRENT APPLICATION NUMBER: US/10/257,909A  
; CURRENT FILING DATE: 2002-10-17  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 3.0

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVGGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVGGFVTAAGITLGMDELYK 239

RESULT 4  
US-10-257-909A-30  
; Sequence 30, Application US/10257909A  
; Publication No. US20030187056A1  
; GENERAL INFORMATION:  
; APPLICANT: Bernard R. TERRY et al.  
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular  
; FILE REFERENCE: 3759-0125P  
; CURRENT APPLICATION NUMBER: US/10/257,909A  
; CURRENT FILING DATE: 2002-10-17  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 30  
; LENGTH: 893  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Fusion between Aequorea victoria and human  
US-10-257-909A-30

Query Match 99.2%; Score 1266; DB 14; Length 893;  
Best Local Similarity 99.6%; Pred. No. 9.8e-123;  
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 60  
DB 655 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYGLTKLFICTTGKLPVPWPT 714  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 715 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 774  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHVMADKQNGIKVNFKIRHNIEDGVSQLA 180  
DB 775 VNRIELKGIDFKEDGNILGHKLEYNYNHVMADKQNGIKVNFKIRHNIEDGVSQLA 834  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVGGFVTAAGITLGMDELYK 239  
DB 835 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVGGFVTAAGITLGMDELYK 893

RESULT 5  
US-10-257-909A-32  
; Sequence 32, Application US/10257909A  
; Publication No. US20030187056A1  
; GENERAL INFORMATION:  
; APPLICANT: Bernard R. TERRY et al.  
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular  
; FILE REFERENCE: 3759-0125P  
; CURRENT APPLICATION NUMBER: US/10/257,909A  
; CURRENT FILING DATE: 2002-10-17  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 3.0

```

; SEQ ID NO 32
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-32

Query Match
Best Local Similarity 99.2%; Score 1266; DB 14; Length 1132;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
DB 894 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 953
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 954 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 1013
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 1014 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 1073
QY 181 DHYQONTPIGDGPVLLPDNHHYLSQSALSKDPNEKRDHNVLGFGVFTAAGITLGMDELYK 239
DB 1074 DHYQONTPIGDGPVLLPDNHHYLSQSALSKDPNEKRDHNVLGFGVFTAAGITLGMDELYK 1132

RESULT 6
US-09-887-784-2
; Sequence 2, Application US/09887784
; Patent No. US20020177189A1
; GENERAL INFORMATION:
; APPLICANT: BJOERN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea Victoria
US-09-887-784-2

Query Match
Best Local Similarity 98.6%; Score 1258; DB 9; Length 239;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHHYLSQSALSKDPNEKRDHNVLGFGVFTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHHYLSQSALSKDPNEKRDHNVLGFGVFTAAGITLGMDELYK 239

RESULT 7
US-10-296-953-2
; Sequence 2, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match
Best Local Similarity 99.2%; Score 1258; DB 12; Length 239;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHHYLSQSALSKDPNEKRDHNVLGFGVFTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHHYLSQSALSKDPNEKRDHNVLGFGVFTAAGITLGMDELYK 239

RESULT 8
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. US20020083488A1
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match
Best Local Similarity 98.4%; Score 1255; DB 9; Length 239;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
```

```

; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match
Best Local Similarity 99.2%; Score 1258; DB 12; Length 239;
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHHYLSQSALSKDPNEKRDHNVLGFGVFTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHHYLSQSALSKDPNEKRDHNVLGFGVFTAAGITLGMDELYK 239

RESULT 8
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. US20020083488A1
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match
Best Local Similarity 98.4%; Score 1255; DB 9; Length 239;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
```



APPLICANT: TSIEN, Roger  
APPLICANT: ZACHARIAS, David  
APPLICANT: BAIRD, Geoffrey  
TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS  
FILE REFERENCE: REGEN1530  
CURRENT APPLICATION NUMBER: US/09/794,308  
CURRENT FILING DATE: 2001-02-26  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
US-09-794-308-4

Query Match 98.4%; Score 1255; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 2.1e-122;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60  
QY 61 LVTLLSYGVCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVGFTVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLEFVTAAGITLGMDELYK 239

RESULT 13  
US-09-865-291-4  
Sequence 4, Application US/09865291  
Publication No. US20030186229A1  
GENERAL INFORMATION:  
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA  
APPLICANT: TSIEN, Roger  
APPLICANT: TING, Alice  
APPLICANT: ZHANG, Jin  
TITLE OF INVENTION: EMISSION RATIO-METRIC INDICATORS OF PHOSPHORYLATION  
FILE REFERENCE: REGEN1550  
CURRENT APPLICATION NUMBER: US/09/865,291  
CURRENT FILING DATE: 2001-05-24  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
US-09-865-291-4

Query Match 98.4%; Score 1255; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 2.1e-122;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60  
QY 61 LVTLLSYGVCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVGFTVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLEFVTAAGITLGMDELYK 239

RESULT 14  
US-10-457-982-3  
Sequence 3, Application US/10457982  
Publication No. US20030212265A1  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger Y.  
APPLICANT: Miyawaki, Atsushi  
APPLICANT: Llopis, Juan  
APPLICANT: Wachter, Rebekka M.  
APPLICANT: Remington, S. James  
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
FILE REFERENCE: 07257/071001  
CURRENT APPLICATION NUMBER: US/10/457,982  
PRIOR FILING DATE: 2003-06-09  
PRIOR APPLICATION NUMBER: US/09/602,641  
PRIOR FILING DATE: 2000-06-22  
PRIOR APPLICATION NUMBER: 09/172,063  
PRIOR FILING DATE: 1998-10-13  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (0)...(0)  
OTHER INFORMATION: EGFP  
US-10-457-982-3

Query Match 98.4%; Score 1255; DB 12; Length 239;  
Best Local Similarity 98.7%; Pred. No. 2.1e-122;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPT 60  
QY 61 LVTLLSYGVCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVGFTVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLEFVTAAGITLGMDELYK 239

RESULT 15  
US-10-121-258-13  
Sequence 13, Application US/10121258  
Publication No. US20030059835A1  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger  
APPLICANT: Campbell, Robert  
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT  
TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME  
FILE REFERENCE: UC083.1CP2CPI  
CURRENT APPLICATION NUMBER: US/10/121,258  
CURRENT FILING DATE: 2002-04-10  
PRIOR APPLICATION NUMBER: 09/794,308  
PRIOR FILING DATE: 2001-02-26  
PRIOR APPLICATION NUMBER: 09/866,538  
PRIOR FILING DATE: 2001-05-24

Sun Jun 27 18:27:30 2004

```
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Enhanced Green Fluorescent Protein (EGFP)
US-10-121-258-13

Query Match      98.4%; Score 1255; DB 14; Length 239;
Best Local Similarity 98.7%; Pred. No. 2.1e-122;
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLT/LKFICTTGKLPVPWPT 60
Db      1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYVKLT/LKFICTTGKLPVPWPT 60

QY      61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKPEGDTL 120
Db      61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKPEGDTL 120

QY      121 VNRLEKIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db      121 VNRLEKIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY      181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239
Db      181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLEFVTAAGITLGMDELYK 239
```

Search completed: June 21, 2004, 16:09:30  
Job time : 34.7778 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:00 ; Search time 10.3333 Seconds  
(without alignments)  
2224.817 Million cell updates/sec

Title: US-09-887-784-222G  
Perfect score: 1276  
Sequence: 1 MVSKEELFTGVVPILVELD.....VLGGFVTAAGITLGMDELYK 239  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1233	96.6	238	1 JQ1514	green-fluorescent
2	105	8.2	785	2 H72228	hypothetical prote
3	92	7.2	461	2 T06936	photosystem II chl
4	91.5	7.2	861	2 T64102	leucine-trna ligas
5	90	7.1	425	2 C97354	hypothetical prote
6	90	7.1	632	2 T06586	DNA-binding protei
7	89.5	7.0	887	2 E82590	leucyl-trna synthe
8	88.5	6.9	655	2 D83917	DNA topoisomerase
9	88	6.9	461	2 S41480	photosystem II chl
10	88	6.9	578	1 I40794	dihydrolipoamide d
11	87.5	6.9	370	2 E70390	iron-sulfur cofact
12	87.5	6.9	860	2 AC0582	leucyl-trna synthe
13	87.5	6.9	2222	1 A36028	DNA-directed DNA p
14	87.5	6.9	2573	2 D71614	hypothetical prote
15	87	6.8	1259	2 AE1055	probable exported
16	86	6.7	357	2 G81355	tRNA (uracil-5)-m
17	86	6.7	877	2 H64708	iron-regulated out
18	85.5	6.7	353	2 E84941	imidazoleglycerol-
19	85.5	6.7	788	1 JDVLHH	DNA-directed DNA p
20	85.5	6.7	889	2 JC5576	inter-alpha-trypsi
21	85.5	6.7	1259	2 H65233	ytfn protein - Esc
22	85.5	6.7	1259	2 G91278	hypothetical prote
23	85.5	6.7	1259	2 G86119	hypothetical prote
24	85	6.7	281	2 AD2052	hypothetical prote
25	85	6.7	459	2 S42647	photosystem II chl
26	85	6.7	459	2 AD2342	photosystem II CP4
27	85	6.7	865	2 T11852	lipoxigenase (sc 1
28	84.5	6.6	613	2 A99552	oligoendopeptidase
29	84.5	6.6	888	2 A54280	cell differentiat

RESULT 1

JQ1514  
green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)  
C;Species: Aequorea victoria  
C;Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 23-Mar-2001  
C;Accession: JS0692; JQ1514; PQ0335; S48693; S51330; S51331  
R;Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.  
Gene 111, 229-233, 1992  
A;Title: Primary structure of the Aequorea victoria green-fluorescent protein.  
A;Reference number: JQ1514; MUID:92175527; PMID:1347277  
A;Accession: JS0692  
A;Molecule type: DNA  
A;Residues: 1-107, 'S', 109-238 <PRA1>  
A;Cross-references: GB:M62654; NID:G155662; PIDN:AAA27722.1; PID:G155663  
A;Accession: JQ1514  
A;Molecule type: mRNA  
A;Residues: 1-99, 'P', 101-140, 'L', 142-218, 'V', 220-238 <PRA2>  
A;Cross-references: GB:M62653; NID:G155660; PIDN:AAA27721.1; PID:G155661  
A;Accession: PQ0335  
A;Molecule type: protein  
A;Residues: 46-64;74-122;132-151;154-183;185-200 <PRA3>  
R;Inouye, S.; Tsuji, F.I.  
FEBS Lett. 351, 211-214, 1994  
A;Title: Evidence for redox forms of the Aequorea green fluorescent protein.  
A;Reference number: S48693; MUID:94364470; PMID:8082767  
A;Accession: S48693  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-24, 'Q', 26-156, 'P', 158-171, 'K', 173-238 <INO>  
A;Cross-references: GB:L29345; NID:G606383; PIDN:AAA58246.1; PID:G606384  
R;Watkins, J.N.; Campbell, A.K.  
Submitted to the EMBL Data Library, January 1995  
A;Reference number: S51330  
A;Accession: S51330  
A;Molecule type: mRNA  
A;Residues: 1-13, 'V', 15-24, 'O', 26-44, 'N', 46-153, 'G', 155-156, 'P', 158-171, 'K', 173-227, 'R', 175-183  
A;Cross-references: EMBL:X83959; NID:G634008; PIDN:CAA58789.1; PID:G634009  
A;Experimental source: clone gfp1  
A;Accession: S51331  
A;Molecule type: mRNA  
A;Residues: 1-24, 'Q', 26-29, 'R', 31-83, 'L', 85-153, 'G', 155-156, 'P', 158-171, 'K', 173-208, 'Q', 175-183  
A;Cross-references: EMBL:X83960; NID:G634010; PIDN:CAA58790.1; PID:G634011  
A;Experimental source: clone gfp2  
R;Yang, F.; Moss, L.G.; Phillips Jr., G.N.  
Submitted to the Brookhaven Protein Data Bank, August 1996  
A;Reference number: A65692; PDB:IGFL  
A;Content: annotation: X-ray crystallography, 1.9 angstroms, residues 'A', 2-79, 'R', 81-91  
A;Note: engineered sequence based on JQ1514, cloned and expressed in Escherichia coli  
R;Yang, F.; Moss, L.G.; Phillips Jr., G.N.  
Nat. Biotechnol. 14, 1246-1251, 1996  
A;Title: The molecular structure of green fluorescent protein.  
A;Reference number: A58953; MUID:98294543; PMID:9631087

glucose dehydrogen  
synergohymenotrophi  
DNA-directed DNA p  
Iga Fc receptor pr  
Iga Fc receptor pr  
hypothetical prote  
transcription regu  
photosystem II chl  
S-layer protein pr  
water-stress-induc  
nitrogenase (EC 1.  
neuraminidase, pro  
protective surface  
probable iron-regu  
hypothetical prote

A:Contents: annotation: X-ray crystallography, 1.9 angstroms  
C:Comment: This protein is excited by the photoprotein aequorin (see PIR-AQJFNV) emitting  
C:Comment: The chromophore of this protein is formed by modification of Ser-dehydro-Tyr-  
C:Genetics:  
A:Gene: GFP  
A:Introns: 69/3; 167/3  
C:Superfamily: green-fluorescent protein  
C:Keywords: chromoprotein; luminescence  
F;65-67/Cross-link: 5-imidazolinone (Ser-Gly) #status experimental  
F;66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match 96.6%; Score 1233; DB 1; Length 238;  
Best Local Similarity 96.6%; Pred. No. 9.1e-96;  
Matches 230; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTKLFCITCTGKLPVWPPTL 61  
DB 1 MSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTKLFCITCTGKLPVWPPTL 60

QY 62 VTTLISGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 121  
DB 61 VTTFISGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDPKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVQLAD 181  
DB 121 NRIELKGIDPKEDGNILGHKWEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVQLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQSALSKDPNKRDMVLGGFVTAAGITLGMDELYK 239  
DB 181 HYQONTPIGDGPVLLPDNHYLSQSALSKDPNKRDMVLLEFVTAAGITHGMDELYK 238

RESULT 2  
H72228  
hypothetical protein TM1624 - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: H72228  
R; Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickley  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.

Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: H72228  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-785 <ARN>  
A:Cross-references: GB:AE001806; GB:AE000512; NID:g4982196; PIDN:AAD36691.1; PID:g498219  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM1624

Query Match 8.2%; Score 105; DB 2; Length 785;  
Best Local Similarity 19.7%; Pred. No. 0.9;  
Matches 45; Conservative 32; Mismatches 71; Indels 84; Gaps 7;

QY 3 SKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTKLFCITCTGKLPVWPPTLV 62  
DB 15 NEGRFSGFEGVPGVQAD-----LVRKGLLPHYPVGM- 46

QY 63 TTLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLVN 122  
DB 47 -----NEDLFKEIDREWIYEREFKEDVKGERVDLVFEGVDTLN 88

QY 123 RIELKGIDPKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVQLADH 182  
DB 89 DVVLNGVYL---GSTEDMFTEYRFDVTNLV-----KEKNLKVYIK-----SPIRVPKT 134

QY 183 YQONTPIGDGPVLLPDNHYLSQSALSKDPNKRDMVLGGFVTAAGITLGM 235  
DB 135 LEQNYGLVGGP-----EDP-----IRGYIRKQAQSYGWD 163

A:Contents: annotation: X-ray crystallography, 1.9 angstroms  
C:Comment: This protein is excited by the photoprotein aequorin (see PIR-AQJFNV) emitting  
C:Comment: The chromophore of this protein is formed by modification of Ser-dehydro-Tyr-  
C:Genetics:  
A:Gene: GFP  
A:Introns: 69/3; 167/3  
C:Superfamily: green-fluorescent protein  
C:Keywords: chromoprotein; luminescence  
F;65-67/Cross-link: 5-imidazolinone (Ser-Gly) #status experimental  
F;66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match 96.6%; Score 1233; DB 1; Length 238;  
Best Local Similarity 96.6%; Pred. No. 9.1e-96;  
Matches 230; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTKLFCITCTGKLPVWPPTL 61  
DB 1 MSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTKLFCITCTGKLPVWPPTL 60

QY 62 VTTLISGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 121  
DB 61 VTTFISGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDPKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVQLAD 181  
DB 121 NRIELKGIDPKEDGNILGHKWEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVQLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQSALSKDPNKRDMVLGGFVTAAGITLGMDELYK 239  
DB 181 HYQONTPIGDGPVLLPDNHYLSQSALSKDPNKRDMVLLEFVTAAGITHGMDELYK 238

RESULT 2  
H72228  
hypothetical protein TM1624 - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: H72228  
R; Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickley  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.

Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: H72228  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-785 <ARN>  
A:Cross-references: GB:AE001806; GB:AE000512; NID:g4982196; PIDN:AAD36691.1; PID:g498219  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM1624

Query Match 8.2%; Score 105; DB 2; Length 785;  
Best Local Similarity 19.7%; Pred. No. 0.9;  
Matches 45; Conservative 32; Mismatches 71; Indels 84; Gaps 7;

QY 3 SKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTKLFCITCTGKLPVWPPTLV 62  
DB 15 NEGRFSGFEGVPGVQAD-----LVRKGLLPHYPVGM- 46

QY 63 TTLSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLVN 122  
DB 47 -----NEDLFKEIDREWIYEREFKEDVKGERVDLVFEGVDTLN 88

QY 123 RIELKGIDPKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVQLADH 182  
DB 89 DVVLNGVYL---GSTEDMFTEYRFDVTNLV-----KEKNLKVYIK-----SPIRVPKT 134

QY 183 YQONTPIGDGPVLLPDNHYLSQSALSKDPNKRDMVLGGFVTAAGITLGM 235  
DB 135 LEQNYGLVGGP-----EDP-----IRGYIRKQAQSYGWD 163

## RESULT 3

T09336  
photosystem II chlorophyll a-binding protein psbC - Cyanophora paradoxa cyanelle  
N;Alternate names: protein CP43  
C:Species: cyanelle Cyanophora paradoxa  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 26-Aug-1999  
C:Accession: T06936  
R;Stewart, V.L.; Michalowski, C.B.; Luffelhardt, W.; Bohnert, H.J.; Bryant, D.A.  
submitted to the EMBL Data Library, July 1995  
A:Description: Nucleotide sequence of the cyanelle genome from Cyanophora paradoxa.  
A:Reference number: Z15840  
A:Accession: T06936  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-461 <STI>  
A:Cross-references: EMBL:U30821; NID:g1016083; PIDN:AAA81279.1; PID:g1016192  
A:Experimental source: strain Pringsheim LB555  
C:Genetics:  
A:Gene: psbC

A:Genome: cyanelle  
A:Superfamily: photosystem II chlorophyll a-binding protein psbC  
C:Keywords: chlorophyll; cyanelle; membrane-associated complex; photosynthesis; photosyst

Query Match 7.2%; Score 92; DB 2; Length 461;  
Best Local Similarity 21.4%; Pred. No. 5.5;  
Matches 58; Conservative 35; Mismatches 88; Indels 90; Gaps 13;

QY 25 GHKFSVSGEGDATYKGLTKLFCITCTGK-----PVWP--TLVTT-----LSYGV--- 69  
DB 208 GDGWIVSDNMEDIIGGHIWLAFCIIGVWHILTKPFWARRALVMSGEAYLSYLAAL 267

QY 70 -----QCFSRYPDHMKQHDFFKSAMPE-GYVQERTIFFKDD-----GN 106  
DB 268 ALMGFIANCFVWFNTAYPSEFPPTGTPPEASQAQATFLVRDORLGANVGSAGPTGLGK 327

QY 107 YKTRA---EVKFEGLTLVNRIELKIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIK 163

DB 328 YLMRSPSGEIIFGGTM-----RFDWTRAPLEPLRGANGLD 364

QY 164 VNFKIRHNIEDGSVQLADHYQONTPIGD---GPVLLPDN---HYLSTQSALSKDPNEKRD 217

DB 365 LT-KIKYDIQPMOERRAAEYMTAPLGSLNSVGVATEINSVNVSPRSWLS-----TS 417

QY 218 HMVLGGFV-----TAAGITLGM 235

DB 418 HFVLGFFLFGHLWHAGRARASGGFEKGLD 448

## RESULT 4

H64102  
leucine-tRNA ligase (EC 6.1.1.4) - Haemophilus influenzae (strain Rd KW20)

N;Alternate names: leucyl-tRNA synthetase

C:Species: Haemophilus influenzae

C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 03-Jun-2002

C:Accession: H64102

R; Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A. J.

; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.

; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghagen, N.S.M.

Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.

A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630; PMID:7542800

A:Accession: H64102

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-861 <TRIG>

A:Cross-references: GB:U32774; GB:L42023; NID:g1573942; PIDN:AAC22581.1; PID:g1573943; T.

C:Genetics:

A:Gene: leuS

C:Superfamily: leucine-tRNA ligase

C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 7.2%; Score 91.5; DB 2; Length 861;  
Best Local Similarity 24.1%; Pred. No. 14;  
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;

QY 50 TTGKLPVPWPTLVTLLSYGVCQSRPDPHMKQDFFKSPAMPEGYQERTIFFKQ-----103  
DB 314 TGDKLPI-WVANFVLMHYGTGAVMAVPAH-DQRDF-----BFAQKYSLPKQVIAPLA 364  
QY 104 DGNKTRAEVKFPGDGLVNRNIELKGDGDKGILGHKLEYNYNHNVYIMADK-QKNGI 162  
DB 365 DEIDLTKQAFVEHGLKLVNSDEFDGKVF--DGAENG-----IADKLEKLG 408  
QY 163 ---KVNFKIRH-----NIBDGSVOLADHYQONTPIGCGVLLPDNHYL- 202  
DB 409 GKQVNVRLRDWGSQRQYWGAPIMPLTLENGDVVPA-----PMEDLPILPEDVWMD 461  
QY 203 STQSALSQDPN 213  
DB 462 GVKSPINADPN 472

RESULT 5  
C97354  
hypothetical protein CAC3702 [imported] - Clostridium acetobutylicum  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C;Accession: C97354  
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: C97354  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-425 <KUR>  
A;Cross-references: GB:AE001437; PIDN:AAK81622.1; PID:g15026806; GSPDB:GN00168  
A;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
A;Gene: CAC3702

Query Match 7.1%; Score 90; DB 2; Length 425;  
Best Local Similarity 24.7%; Pred. No. 7.3;  
Matches 66; Conservative 31; Mismatches 98; Indels 72; Gaps 16;

QY 21 GDVNGHKFSVSGEGDATYCKLTKEICTTK-----LVPWPTLVTLLSYGVC- 71  
DB 109 GFVNGKLLPASGEAIKD--FGKV----LKNTGDKLGLTRQVEMFPG-VGRVSVGPNTIE 161  
QY 72 FSRYPDHMK-----QHDFFKSPAMPEGYQERTIFFKDDGNKYTR-----AEVKFEGD-- 118  
DB 162 FSEVXNFKNGVGDIKDNFVAKVSEDKVMGNP---SGEGNYTEWYNLYKKEKYGEDNV 218

QY 119 -----TLVNRNIELKIDPK-----DGNILGHKLEYN-----NSH--NVYIMA 155  
DB 219 YLCTDEKSLINSVSEKLSFRKNGVDVYG-KTAVAYGENLDTVSMATNSHQYNPILT 277

QY 156 DKQKNGIKVNFKIRHIEDSGVOLADHYQONTPIGCGVLLPDNHYLSTQSALSKD---- 211  
DB 278 DGEKFDKFNMYNKLKEYGDESVAISKY--BHTLTEDYIKTLPDEYF--TSKGLVKKDDVIS 334

QY 212 -----PNEKRDMVLGGFVTAAG 229  
DB 335 EMKKLTDALIENTNREARIAGKFKSLTG 361

RESULT 6  
T06586  
DNA-binding protein PD2 - garden pea  
C;Species: Pisum sativum (garden pea)  
C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 08-Oct-1999  
C;Accession: T06586  
R;Sato, N.; Kazuno, A.A.; Ohta, N.; Oshihama, K.

submitted to the EMBL Data Library, June 1996  
A;Description: Identification of a novel family of DNA-binding proteins with two AT-hook  
A;Reference number: Z15774  
A;Accession: T06586  
A;Status: preliminary; translated from GB/EMBL/DBSJ  
A;Molecule type: mRNA  
A;Residues: 1-632 <SAT>  
A;Cross-references: EMBL:X98740; NID:e9952239; PIDN:CAA67292.1; PID:e275185  
A;Experimental source: cv. Alaska

Query Match 7.1%; Score 90; DB 2; Length 632;  
Best Local Similarity 23.3%; Pred. No. 12;  
Matches 49; Conservative 26; Mismatches 79; Indels 56; Gaps 7;

QY 16 LVLEDGDVNGHKFSVSGEGDATYCKLTKEICTTKLPVWPTLVTLLSYGVCFSRY 75  
DB 363 IVDOGRDVGSKVDVINKESNEATIPENK-----PTEPKLDVEQELAATM----- 408  
QY 76 PDHMKQDFFKSPAMPEGYV-----QERTIFFKDDGNKYKTRAEVKFEGDTLVNRIE 125  
DB 409 PSSAKVNVLTLDLIVETFPILRSVARTSSGREGSEELKDSGNSLERDTKKLEBQGNKNS-E 467  
QY 126 LKGIKDEKDGNIHLGHKLE-----YVNSHNVYIMADKQK-----N 160  
DB 468 LKGIETDSTLLDEKFNALGNKILKEISNPRHDVESANHSHTNKQVTVSHQKAIETNN 527  
QY 161 GIKVNFKIRHNIEDG-----SVQLADHYQ 184  
DB 528 QSQVEDVAKNKIQDDSKPSESLHKADKYR 557

RESULT 7  
E82590  
leucyl-tRNA synthetase XF2176 [imported] - Xylella fastidiosa (strain 9a5c)  
C;Species: Xylella fastidiosa  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
C;Accession: E82590  
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequence  
Nature 406, 151-157, 2000  
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A;Reference number: A82515; MUID:20365717; PMID:10910347  
A;Note: for a complete list of authors see reference number A59328 below  
A;Accession: E82590  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-887 <SIM>  
A;Cross-references: GB:AE004031; GB:AE003849; NID:g9107309; PIDN:AAF84975.1; GSPDB:GN00162  
A;Experimental source: strain 9a5c  
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; AJ  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H  
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Praga, J.S.; Franca, S.C.; Franco, M.C.; Frohme  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigret  
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaka  
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira  
M.; Teshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A;Reference number: A59328  
A;Contents: annotation  
C;Genetics:  
A;Gene: XF2176  
C;Superfamily: leucine-tRNA ligase

Query Match 7.0%; Score 89.5; DB 2; Length 887;  
Best Local Similarity 22.2%; Pred. No. 21;  
Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;

QY 50 TTGKLPVPWPTLVTLLSYGVCQSRPDPHMKQDFFKSPAMPEGYQERTIFFKDDGNY-- 107  
DB 329 TNEQLPV-WVANFVLMAYGTGAVMAVPGHQDQRQDEF--ANKYGLPIRQVIALKEPKNQDE 385

QY 108 -----KTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHVYI 153  
DB 386 STWEPDVRWDYADKTR---EPF---LINSAPEDGLDYQDAFEVLAERFE----- 429  
QY 154 MADKQKNG-IKNFKIRHNIEDGSVQLADHYQOQNTPI-----GDGPVLLPDN 199  
DB 430 ---RQGRQRRVNYRLR----DWGVSQRVYWGCPPIVYICPTCGAVVPEDQLPVILPEN 482  
QY 200 -HYLSTQSALSADPNNEKR 216  
DB 483 VAPSGTGSPIKTDPWRK 500  
RESULT 8  
DB3917  
DNA topoisomerase IV subunit B BH2140 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C:Accession: D83917  
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000  
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A;Reference number: A83650; MUID:20512582; PMID:11058132  
A;Accession: D83917  
A;Status: Preliminary  
A;Molecule type: DNA  
A;Residues: 1-655 <STO>  
A;Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BAB05859.1; GSPDB:GN00  
A;Experimental source: strain C-125  
C:Genetics:  
A;Gene: BH2140  
C;Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B  
Query Match 6.9%; Score 88.5; DB 2; Length 655;  
Best Local Similarity 21.9%; Pred. No. 17;  
Matches 53; Conservative 37; Mismatches 95; Indels 57; Gaps 9;  
QY 22 DVNGHK---FSVSGEGEGDAT---YKLTLLKFI-----CTTGKLPVPWP 59  
DB 63 NVTHDKQSVSRDEGRGMPGMHKLKGPTEPVILTVLHAGGKFGGGVATSGGLHGVGA 122  
QY 60 TLVTTLSSYGQCFSPYDPMKQHDFFKSAMPEGYQVER-----TIFFKDDG----- 105  
DB 123 SVVNALSEWLIVKIDGHWYQRFENGKGPSTLEKKGKTRGTGTHFKPDPVFSIT 182  
QY 106 --NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADK----- 157  
DB 183 NFNVETLSERLEAAFLKGLKIELVLDLDDTKVEFH-YEDGKAFVEYLNEDKETLHPV 241  
QY 158 -----QKNGIKVNFKIRHNIEDGSVQLADHYQOQNTPIGDGPVLLPDNHYLSTQSALSADPN 212  
DB 242 VFPNGESNGIEIFAFQFN--DGYTENVLVSFVNVRTKDG-----GTHELGAKTAMTRAV 294  
QY 213 NE 214  
DB 295 NE 296  
RESULT 9  
S41480  
Photosystem II chlorophyll a-binding protein psbC - Chlamydomonas eugametos chloroplast  
C:Species: chloroplast Chlamydomonas eugametos  
C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 26-Aug-1999  
C:Accession: S41480  
R;Cote, M.J.J.; Lemieux, C.; Turmel, M.  
submitted to the EMBL Data Library, May 1993  
A;Description: Identification of self-splicing group I introns in the chloroplast psbC  
of Chlamydomonas moewusii.  
A;Reference number: S41480  
A;Accession: S41480  
A;Molecule type: DNA  
A;Residues: 1-461 <COT>

A;Cross-references: EMBL:M90639; NID:g289948; PIDN:AAA84148.1; PID:g289949  
C:Genetics:  
A;Gene: psbC  
A;Genome: Chloroplast  
A;Start codon: GTG  
A;Introns: 181/3; 294/3  
C;Superfamily: photosystem II chlorophyll a-binding protein psbC  
C;Keywords: chlorophyll; chloroplast; membrane-associated complex; photosynthesis; photos

Query Match 6.9%; Score 88; DB 2; Length 461;  
Best Local Similarity 19.5%; Pred. No. 12;  
Matches 51; Conservative 37; Mismatches 83; Indels 90; Gaps 12;

QY 37 DATYKGLTLKFICTTGKL---PVPWP-----TLVTTLSSYGQCFSR 74  
DB 220 DIIGHIWTGTCIFGCIWHIYTPWPWARRAFVMSGEAYLSVSLGAIATMGFIACFSW 279  
QY 75 YPDHMKQHDFFKSAMPE-GYQERTIFFKDD-----GNKYTRA---EVKF 115  
DB 280 FNNATVPSEFYGPTGPEASQAQAFTLVRDQRLGANVASAQGPTGLGKYLMSRPTGEIIF 339  
QY 116 EGDTLVNRTELKIDFKEDGNILGHKLEYNNSHVYIMADKQKGIKVNFKIRHNIEDG 175  
DB 340 GGET-----MRFWDFR-----GPWCPLRGPNGLDLN-KLUNDIQPW 375  
QY 176 SVQLADHYQOQNTPIGD---GPVLLPDN--HYLSTQSALSADPNNEKRDMHVLGGF----- 224  
DB 376 QERRAAEYTHAPLGLSLNSVGGVATEINAVNFVSPRWLA-----TSHFVLGFFFFVGH 429  
QY 225 -----VTAAGITLGMDEL 237  
DB 430 LWHAGRARAAAAGFEKGIDRV 450

## RESULT 10

I40794

dihydroliipoamide dehydrogenase (EC 1.8.1.4) [validated] - Clostridium magnum  
N;Alternate names: 2-oxoglutarate dehydrogenase complex chain E3; acetoacetyl dehydrogenase  
hydrogenase complex chain E3; S-complex 50K chain  
C;Species: Clostridium magnum  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
C;Accession: I40794  
R;Kruger, N.; Oppermann, F.B.; Lorenzl, H.; Steinbuchel, A.  
J. Bacteriol. 176, 3614-3630, 1994  
A;Title: Biochemical and molecular characterization of the Clostridium magnum acetoacetyl de  
A;Reference number: I40789; MUID:94266715; PMID:8206840  
A;Accession: I40794  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-578 <KRU>  
A;Cross-references: GB:L31844; NID:g472324; PIDN:AAA21748.1; PID:g472330  
C;Function:  
A;Description: catalyzes the oxidation of dihydroliipoamide to liipoamide using NAD  
A;Pathway: acetoacetyl dehydrogenase enzyme system  
C;Superfamily: Alkaligenes dihydroliipoamide dehydrogenase; dihydroliipoamide dehydrogenase  
C;Keywords: FAD; flavoprotein; liipoamide; NAD; oxidoreductase; redox-active disulfide  
F;5-77/Domain: lipoyl/biotin-binding homology <LPS>  
F;117-145/Region: beta-alpha-beta FAD nucleotide-binding fold  
F;119-561/Domain: dihydroliipoamide dehydrogenase homology <DLN>  
F;287-315/Region: beta-alpha-beta NAD nucleotide-binding fold  
F;153-158/Disulfide bonds: redox-active #status predicted

Query Match 6.9%; Score 88; DB 1; Length 578;  
Best Local Similarity 22.9%; Pred. No. 16;  
Matches 57; Conservative 43; Mismatches 91; Indels 58; Gaps 13;

QY 10 TGVVPIVLVDGNGHKSVSSEGEGDATYKLTUK-----FICTGKLPWPPTLVT 63  
DB 255 TGSMPPIPIEIE---GNKLS---GVIDST-GALSLESPESIAIIGGGVIGVEFASIFN 305  
QY 64 TLSYGQCFSPYDPMKQHDFFKSAMPEGYQVERTIFFKDDGNYKTRAEVKFEGDTLVNR 123  
DB 306 SLGCKSVIIEMLPHILPMPMDREISEI-----AKALIRGININN 346

A/Accession: S65121  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Molecule: 1-2231 <SEF>  
A/Residues: 1-2231  
A/Cross-references: EMBL:G1045236; NID:G1045236; PID:CAA63235.1; PID:G1045247  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1982  
C/Genetics:  
C/Gene: SGD: YNL262W  
C/Gen: MIPS: YNL262W

RESULT 15  
AE1055  
probable exported protein ytfN [imported] - Salmonella enterica subsp. enterica serovar  
C;Species: Salmonella enterica subsp. enterica serovar typhi  
A;Note: this species has also been called Salmonella typhi  
C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C;Accession: AE1055  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
th, T.; Cronin, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A;Reference number: AB0502; PMID:21534947; PMID:11677608

Search completed: June 21, 2004, 16:02:00  
Job time : 11.3333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:42:24 ; Search time 6.44444 Seconds  
(without alignments)  
1931.085 Million cell updates/sec

Title: US-09-887-784-222G

Perfect score: 1276

Sequence: 1 MVSKEELFTGVVPIIVELD.....VLGFFVTRAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1239	97.1	238	1	GFP_AEQVI
2	92	7.2	461	1	PSBC_CYPAPA
3	91.5	7.2	861	1	SYL_HAEIN
4	89.5	7.0	879	1	SYL_XYLFA
5	88	6.9	461	1	PSBC_CHLEU
6	87.5	6.9	860	1	SYL_SALTI
7	87.5	6.9	860	1	SYL_SALTY
8	87.5	6.9	222	1	DPOE_YEAST
9	87	6.8	689	1	AC2L_HUMAN
10	86.5	6.8	533	1	CP51_CANGA
11	86.5	6.8	879	1	SYL_XYLFT
12	86.5	6.8	1603	1	VITI_CABEL
13	86	6.7	357	1	TRMA_CAMJE
14	85.5	6.7	353	1	HIS7_BUCAI
15	85.5	6.7	658	1	ADAS_HUMAN
16	85.5	6.7	788	1	DPOL_HPBHE
17	85.5	6.7	886	1	ITH3_MSAU
18	85.5	6.7	1259	1	YTFN_ECOLI
19	85	6.7	504	1	YC03_KLEPN
20	84.5	6.6	613	1	PEPF_MYCPU
21	84.5	6.6	859	1	SYL_SHEON
22	84.5	6.6	888	1	LAGG_DICDI
23	84	6.6	366	1	SET7_HUMAN
24	83.5	6.5	538	1	GRBE_RAT
25	83.5	6.5	589	1	SYD_HAEDU
26	83.5	6.5	1164	1	BAG_STRAG
27	83	6.5	472	1	PSBC_SYNY3
28	83	6.5	874	1	SLAP_BACLI
29	82.5	6.5	501	1	AMPA_WTGBR
30	82.5	6.5	533	1	NIFD_CLOPA
31	82.5	6.5	658	1	ADAS_CAVPO
32	82.5	6.5	795	1	D152_HAEIN
33	82.5	6.5	795	1	D151_HAEIN

34	82	6.4	682	1	PRC_ECOLI	P23865	escherichia
35	82	6.4	728	1	CATB_ASPFU	Q92405	aspergillus
36	82	6.4	752	1	NEC1_RAT	P28840	rattus norv
37	81.5	6.4	793	1	D153_HAEIN	O32629	haemophilus
38	81.5	6.4	941	1	GUN_BACS6	P19424	bacillus sp
39	81	6.3	336	1	YD48_METJA	O58743	methanococ
40	81	6.3	682	1	AC2L_MOUSE	Q99nb1	mus muscucu
41	81	6.3	737	1	OPT1_DROME	P91679	drosophila
42	81	6.3	774	1	AMY2_SCHPO	O42918	schizosacch
43	80.5	6.3	393	1	TRMB_HELPY	O25443	helicobacte
44	80.5	6.3	860	1	SYL_ECO57	Q8xbn8	escherichia
45	80.5	6.3	860	1	SYL_ECOL6	Q8fjy9	escherichia

ALIGNMENTS

RESULT 1

GFP\_AEQVI STANDARD; PRT; 238 AA.

ID AC P4212; Q17104; Q27903;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Green fluorescent protein.

GN GFP.

OS Aequorea victoria (Jellyfish).

OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;

OC Aequoreidae; Aequorea.

OX NCBI\_TaxID=6100;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RX MEDLINE=92175527; PubMed=1347277;

RA Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,

RA Cormier M.J.;

RT "Primary structure of the Aequorea victoria green-fluorescent

RT protein.";

RL Gene 111:229-233(1992).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=94185810; PubMed=8137953;

RA Inouye S., Tsuji F.I.;

RT "Aequorea green fluorescent protein. Expression of the gene and

RT fluorescence characteristics of the recombinant protein.";

RL FEBS Lett. 341:277-280(1994).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=97299832; PubMed=9154981;

RA Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;

RT "Enhanced expression in tobacco of the gene encoding green fluorescent

RT protein by modification of its codon usage.";

RL Plant Mol. Biol. 33:989-999(1997).

RN [4]

RP CHROMOPHORE.

RX MEDLINE=93192221; PubMed=8448132;

RA Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;

RT "Chemical structure of the hexapeptide chromophore of the Aequorea

RT green-fluorescent protein.";

RL Biochemistry 32:1212-1218(1993).

RN [5]

RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).

RX MEDLINE=96355665; PubMed=8703075;

RA Ormoe M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,

RA Remington S.J.;

RT "Crystal structure of the Aequorea victoria green fluorescent

RT protein.";

RL Science 273:1392-1395(1996).

RN [6]

RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).

RX MEDLINE=98294543; PubMed=9631087;

RA Yang F., Moss L.G., Phillips G.N. Jr.;

RT "The molecular structure of green fluorescent protein.";

RL Nat. Biotechnol. 14:1246-1251(1996).





```
Db 181 HYQONTPIGDPVLLPDNHYLSQTSALSXDPNEKRDHMLLEFVTAAGITHGMDELYK 238
RESULT 2
PSBC_CYPAPA STANDARD; PRT; 461 AA.
AC P48104;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Photosystem II 44 kDa reaction center protein (P6 protein) (CP43).
GN PSBC.
OS Cyanophora paradoxa.
OG Cyanelle.
OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OX NCBI_TaxID=2762;
[1]
SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
RA Bryant D.A.;
RA "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa.";
RL Plant Mol. Biol. Rep. 13:327-332(1995).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
RA Farley J.Y., Schluchter W.M., Chung S., Newmann-Spallart C.,
RA Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
RT the genetic complexity of a primitive plastid.";
RL (in) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,
RL Schwemmler W. (eds.);
RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
RL (1997).
CC -!- FUNCTION: The 43 kDa protein (p6) is a component of the core of
CC photosystem II. It is a chlorophyll binding protein.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Cyanelle
CC thylakoid membrane.
CC -!- SIMILARITY: Belongs to the psbB / psbC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U30821; AAA81279.1; -.
CC PIR; T06936; T06936
CC InterPro; IPR005849; Photo44.
CC InterPro; IPR000932; PSI1prot.
CC Pfam; PF00421; PSII; 1.
CC TIGRFAMs; TIGR01153; psbc; 1.
CC Photosynthesis; Photosystem II; Thylakoid; Chlorophyll; Cyanelle;
CC Transmembrane.
CC SEQUENCE 461 AA; 50402 MW; 6264E3239E13B9C1 CRC64;
Query Match 7.2%; Score 92; DB 1; Length 461;
Best Local Similarity 21.4%; Pred. No. 3;
Matches 58; Conservative 35; Mismatches 88; Indels 90; Gaps 13;
QY 25 GHKFSVSGEGDATYKGLTKLICTTGK-----PVPWP--TLVTT---LSYGV--- 69
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 GDGWISVDNMEDIIGGHIWLAFLICIGGWHILTKPFSWARRALVWSGEAVLSYLAAL 267
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 70 -----QCFSRYPDHMKQHDFFKSAMPE-GYVQERTIFPKD-----GN 106
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 268 ALMGFIANCVFVFNNTAYSEFFGPTGPEASQAQFTFLVRDQRLGANVGSAGPTGLGK 327
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 107 YKTRA---EVKFFGDTLVNRNIELKGIDFKEDGNILGHKLEYNNSHNVIMADKQKNGIK 163
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 328 YLMRSPSGBIIFGGETM-----RFWDTRAPWLEPLRGANGLD 364
164 VNFKIRHNIEDSGVOLADHYQONTPIGD---GPVLLPDN---HYLSQTSALSXDPNEKRD 217
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 365 LT-KIKYIQPWQERRAAEYMTAPLGLSLNSVGGVATEINSVYVSPRWLS-----TS 417
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 218 HMVLGGFV-----TAAGITLGM 235
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 418 HFVLGFFLFIQHLWHAGRARAAASGGFEGKGLD 448
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 3
SYL_HAEIN STANDARD; PRT; 861 AA.
AC P43827;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR HI0921.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RC MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McInnes K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32774; AAC22581.1; -.
CC PIR; H64102; H64102.
CC TIGR; HI0921; -.
CC HAMAP; MF_00049; -.
CC InterPro; IPR002302; Leu-tRNA-synthla.
CC InterPro; IPR002300; tRNA-synt_1a.
CC InterPro; IPR001412; tRNA-synt_1.
CC InterPro; IPR009008; ValRS_1leRS_edit.
CC Pfam; PF00133; tRNA-synt_1_1.
CC PRINTS; PRO0985; TRNASYNTHLEU.
CC TIGRFAMs; TIGR00396; leuS_bact; 1.
CC PROSITE; PS00178; AA tRNA LIGASE 1; 1.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
CC SEQUENCE 861 AA; 97750 MW; EB93304F6B4C8FB7 CRC64;
```

```
Query Match 7.2%; Score 91.5; DB 1; Length 861;
Best Local Similarity 24.1%; Pred. No. 7;
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;

QY 50 TTGKLPVWPPTLTVTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKRD-----EFAQKYSIPKQVIAPLA 364
DB 314 TGDKLPV-VWVNFVLMHYGTGAVMAPAH-DQRDF-----EFAQKYSIPKQVIAPLA 364

QY 104 DGNKYTRAEVKFGDGLVNRVIEKLGIDFKEDGNILGHKLEYNVNSHNVMADK-QKNGI 162
DB 365 DEBIDLTKQAFVHGKLVNSDEPDGKNF--DGAFFNG-----IADKLEKLG 408

QY 163 ---KNFKIRH-----NIEDSGVQLADHYQQNTPIGDPVLLPDNHYL- 202
DB 409 GKQVNYRLDVGVSQRVYWGAPIMLTLENGDVVFA-----PWEDLPILPEDVVWD 461

QY 203 STQSALSADPN 213
DB 462 GVKSPINADPN 472

RESULT 4
SYL_XYLFA STANDARD; PRT; 879 AA.
AC Q9PEG8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (SC 6.1.1.4) (Leucine--tRNA ligase) (Leurs).
GN LEUS CR XF2176.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RC STRAIN=9a5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barro M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Coutinho L.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.B., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.P., Truffi D., Teai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Secubal J.C.;
RA "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diposphate + L-leucyl-tRNA(Leu).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AE004031; AAF84975.1; ALT_INIT.
CC HAMAP; MF_00049; -; 1.
CC InterPro; IPR002302; Leu-trnasyntla.
CC InterPro; IPR002300; tRNA-synt_1a.
CC InterPro; IPR001412; tRNA-synt_1.
CC InterPro; IPR009008; Valrs_1fiers_edit.
CC Pfam; PF00133; tRNA-synt_1; 1.
CC PRINTS; PR00985; TRNASYNTHLEU.
CC TIGRFAMs; TIGR00396; leus_bact; 1.
CC PROSITE; PS00178; AA_TRNA_LIGASE; 1.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC SITE 45 55 "HIGH" REGION.
CC SITE 637 641 "KMSKS" REGION.
CC BINDING 640 640 ATP (BY SIMILARITY).
CC SEQUENCE 879 AA; 9796 MW; 9FDCB992092919E CRC64;
CC
CC Query Match 7.0%; Score 89.5; DB 1; Length 879;
CC Best Local Similarity 22.2%; Pred. No. 10;
CC Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;
CC
CC QY 50 TTGKLPVWPPTLTVTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDGNY-- 107
CC DB 321 TNEQLFPV-VWVNFVLMHYGTGAVMAPVGHQDRQDF--ANKYGLPIRQVIALKEPKNQDE 377
CC QY 108 -----KTRAEVKFEGDTLVNRIELKIDGIDFKEDGNILGHKLEYNVNSHNVI 153
CC DB 378 STWEPDVWRDWDYADKTR---EFE---LINSAFDGLDYQDAFEVLAERFE----- 421
CC QY 154 MADKQKNG-IKYNFKIRHNIEDSGVQLADHYQQNTPI-----GGGVLLPDN 199
CC DB 422 ---RQGRQRRVNYRLR----DWGVSQRVYWGCPFVIYCTCGAVPVPEDQLPVILPEN 474
CC QY 200 -HYLSQTQSALSADPN 216
CC DB 475 VAFSGTGSPIKTDPEWRK 492
CC
CC RESULT 5
CC PSBC_CHLEU STANDARD; PRT; 461 AA.
CC AC Q08684;
CC DT 01-OCT-1994 (Rel. 30, Created)
CC DT 01-OCT-1994 (Rel. 30, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Photosystem II 44 kDa reaction center protein (P6 protein) (CP43).
CC GN PSBC.
CC OS Chlamydomonas eugametos.
CC OC Chloroplast.
CC OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
CC OC Chlamydomonadaceae; Chlamydomonas.
CC OX NCBI_TaxID=3053;
CC [1]
CC SEQUENCE FROM N.A.
CC MEDLINE=9407731; PubMed=7504814;
CC RX Turmel M., Mercier J.P., Cote M.J.J.;
CC "Group I introns interrupt the chloroplast psbA and psbC and the
CC RT mitochondrial rnl gene in Chlamydomonas.";
CC Nucleic Acids Res. 21:5242-5250(1993).
CC -1- FUNCTION: The 43 kDa protein (p6) is a component of the core of
CC photosystem II. It is a chlorophyll binding protein.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast
CC thylakoid membrane.
CC -1- SIMILARITY: Belongs to the psbB / psbC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC
```



RN Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RP Waterston R., Wilson R.K.;  
RC "Complete genome sequence of Salmonella enterica serovar Typhimurium  
LT2.";  
RA Nature; 413:852-856(2001).  
RT -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +  
CC diphosphate + L-leucyl-tRNA(Leu);  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: Belongs to class-I aminocacyl-trna synthetase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to licens@isb-sib.ch).

EMBL; AE008725; AAL19599.1; --  
STyGene; SG?????; leus.  
HAMAP; MF\_00049; -; I.  
InterPro; IPRO02302; Leu-trnasyntla.  
InterPro; IPRO02300; trna-synt la.  
InterPro; IPRO01412; trna-synt l.  
InterPro; IPRO09008; ValRS\_IleRS\_edit.  
Pfam; PF00133; trna-synt\_l; 1.  
PRINTS; PR00985; TRNASYNTHLEU.  
TIGRFAMs; TIGR00396; leus\_bact; 1.  
DR PROSITE; PS00178; AA TRNA LIGASE I; 1.  
KW Aminocacyl-trna synthetase; Protein biosynthesis; Ligase; ATP-binding;

KW Complete proteome.

FT SITE 42 52 "HIGH" REGION.  
FT SITE 619 623 "KMSKS" REGION.  
FT BINDING 622 622 ATP (BY SIMILARITY).  
SQ SEQUENCE 860 AA; 96985 MW; D5003584DFBCAB6 CRC64;

Query Match 6.9%; Score 87.5; DB 1; Length 860;  
Best Local Similarity 23.3%; Pred. No. 15;  
Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;

QY 50 TTGKLVPVMTLVTLTSYGVCFSRYPDHMQKHDPFKSAMPEGVVQERTIFFKDGNKYKT 109  
DB :|||:  
DQ 314 TGEEIPV-WAANFVLMEYTGAVMAVGPH-DQRD-YEFASKYLGTIRPVILAADGSPPDL 370  
QY |||:  
DQ 110 RAEVKEFGDTLNRLIELKGIDKFEDGNILGHKLYNNHNVMADMCKNGIKVNPKIR 169  
DB :|||:  
DQ 371 SEQALTGKVLFNSGEFDGLAEAFANAIDL-----AEKGVERKNVYELR 418  
QY :|||:  
DQ 170 H-----NIEDSGVOLADHYQQNTPIGDGPVLLPNHV-L-STQSALSNDP 212  
DB :|||:  
DQ 419 DMGVSRQRYGAPIPVTLEDGTIV-----LPTPEDQLPVLPEDVMDGITSPISKADP 471

RESULT 8  
DPOE YEAST STANDARD; PRT; 2222 AA.  
AC AC P21951;  
DT 01-AUG-1991 (Rel. 19, Created)  
DD 01-AUG-1991 (Rel. 19, Last sequence update)  
DI 10-OCT-2003 (Rel. 42, Last annotation update)

DNA polymerase epsilon, catalytic subunit A (EC 2.7.7.7) (DNA polymerase II subunit A).  
DE POL2 OR DUN2 OR YNL262M OR N0825.  
DS Saccharomyces cerevisiae (Baker's yeast).  
OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycosida.  
OX NCBI\_TaxID=4932;  
[1]  
RN SEQUENCE FROM N.A., AND SEQUENCE OF 1214-1221.  
RP MEDLINE=90381771; PubMed=2169349;  
RA Morrison A., Araki H., Clark A.B., Hamatake R.K., Sugino A.;  
RT "A third essential DNA polymerase in S. cerevisiae."; Cell 62:1143-1151(1990).

[2] SEQUENCE OF 1-2221 FROM N.A.  
RC STRAIN=S288c / FYI679;  
RA MEDLINE=96310631; PubMed=8740425;  
RX Sen-Gupta M., Lyck R., Fleig U., Niedenthal R.K., Hegemann J.H.;  
RT "The sequence of a 24,152 bp segment from the left arm of chromosome XIV from Saccharomyces cerevisiae between the BN11 and the POL2 genes.";  
RT Yeast 12:505-514 (1996).

[3] TEMPERATURE SENSITIVE MUTANTS.  
RP MEDLINE=92164663; PubMed=1537345;  
RA Araki H., Ropp P.A., Johnson A.L., Johnston L.H., Morrison A., Sugino A.;  
RT "DNA Polymerase II, the probable homolog of mammalian DNA polymerase epsilon, replicates chromosomal DNA in the yeast Saccharomyces cerevisiae";  
RT EMBO J. 11:733-740 (1992).

-!- FUNCTION: DNA POLYMERASE II PARTICIPATES IN CHROMOSOMAL DNA REPLICATION.  
CC -!- [DNA](N).  
CC KDA, AND 29 kDa).  
CC -!- SUBUNIT: CONSISTS OF FIVE SUBUNITS (200 kda, 80 kda, 34 kda, 30 kda, AND 29 kda).  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- DOMAIN: THE DNA POLYMERASE ACTIVITY DOMAIN RESIDES IN THE N-TERMINAL HALF OF THE PROTEIN, WHILE THE C-TERMINUS IS NECESSARY FOR COMPLEXING SUBUNITS B AND C.  
CC -!- MISCELLANEOUS: In eukaryotes there are five DNA polymerases: alpha, beta, gamma, delta, and epsilon which are responsible for different reactions of DNA synthesis.  
CC -!- SIMILARITY: Belongs to the DNA polymerase type-B family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement or send an email to licens@isb-sib.ch).

EMBL; M60416; AAA88711.1; --  
EMBL; X92494; CAAG3235.1; --  
EMBL; Z71538; CAA96169.1; --  
PIR; A36028; A36028.  
GerOnline; 143268; --  
SGD; S0005206; POL2.  
GO; GO:000731; P:DNA repair synthesis; IMP.  
InterPro; IPR006172; DNA pol\_B.  
InterPro; IPR006134; DNA pol\_B\_dom.  
InterPro; IPR006133; DNA pol\_B\_exo.  
Pfam; PF00136; DNA\_pol\_B; 1.  
Pfam; PF03104; DNA pol\_B\_exo; 1.  
SMART; SM00486; POLBc; 1.  
PROSITE; PS00116; DNA POLYMERASE B; FALSE NEG.  
Transferase; DNA-directed RNA polymerase; DNA replication;  
Transferring; zinc-finger; Nuclear protein.  
ZNFING 2108 2181 POTENTIAL.  
FT VARIANT 644 644 M->I (IN POL2-9 TS MUTANT).  
FT VARIANT 710 710 P->S (IN POL2-18 TS MUTANT).  
SQ SEQUENCE 2222 AA; 255669 MW; CBCDDE2AB147D65B CRC64;

Query Match 6.9%; Score 87.5; DB 1; Length 2222;  
Best Local Similarity 28.2%; Pred. No. 46;  
Matches 37; Conservative 14; Mismatches 49; Indels 31; Gaps 7;

QY 54 LPVPWP-TVTVTTLSYGVCFSYPDHM-----KOHDFFKSAMPGYV---QEITI 99  
DB |||:  
DQ 883 LPKSPETVFVTFLENGKKLYSLPSCLNYRVHQFTTNHOVELKDPLNYIETHSENTI 942  
QY |||:  
DQ 100 FKDDGNKYTR--AEVKEGDTLVNR-----IELGKDFDKDGNGILGHKLEYNN 147  
DB |||:  
DQ 943 FFVDGGPYKAMILPSRKEEKGIKKRYAVFNEDSGLAELKGFELKRREL----QLINKFO 999

QY 148 S--HNVYIMAD 156  
| : : :  
Db 1000 SDIFKVFLECD 1010

RESULT 9  
AC2L HUMAN STANDARD; PRT; 689 AA.  
ID QSNUB1: Q81V99; Q96J11; Q96JX6; Q9NU28;  
AC QSNUB1: Q81V99; Q96J11; Q96JX6; Q9NU28;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Acetyl-coenzyme A synthetase 2-like, mitochondrial precursor  
DE (EC 6.2.1.1) (Acetyl-CoA ligase 2) (Acetyl-CoA synthetase 2)  
GN ACAS2L OR KIAA1846  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21638749; PubMed=11780052;  
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,  
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
RA Coulson A., Coville G.J., Deadman R., Dhali P.D., Dunn M.,  
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
RA Lehesvahti M.H., Leverhwa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
RA Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.A.,  
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.L.,  
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,  
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,  
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
RA Rogers J.;  
RL "The DNA sequence and comparative analysis of human chromosome 20.";  
RN Nature 414:665-871(2001).  
RP [2]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RC TISSUE=Brain, and Testis;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Viallon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences.";

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[3]  
RP SEQUENCE OF 169-689 FROM N.A. (ISOFORM 1).  
RC TISSUE=Placenta, and Tongue;  
RX MEDLINE=21245130; PubMed=11347906;  
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XX.  
The complete sequences of 100 new cDNA clones from brain which code  
for large proteins in vitro.";  
RL DNA Res. 8:85-95(2001).  
CC -!- FUNCTION: Converts acetate to acetyl-CoA so that it can be used  
for oxidation through the tricarboxylic cycle to produce ATP and  
CO(2) (By similarity).  
CC -!- CATALYTIC ACTIVITY: ATP + acetate + CoA = AMP + diphosphate +  
acetyl-CoA.  
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix (By similarity).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC IsoId=Q9NUB1-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q9NUB1-2; Sequence=VSP\_007249;  
CC Notes=No experimental confirmation available;  
CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme  
family.  
CC -!- CAUTION: Ref.1 (CAB81884) sequence differs from that shown due to  
erroneous gene model prediction.  
CC -!- CAUTION: Ref.3 sequence differs from that shown due to a  
frameshift in position 250 and numerous sequencing errors.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AL035661; CAB75500.1; -  
DR EMBL; AL080312; CAB81884.1; ALT\_SEQ.  
DR EMBL; BC039261; AAH39261.1; -  
DR EMBL; BC044588; AAH44588.1; -  
DR EMBL; AK027817; BAB55390.1; ALT\_INIT.  
DR EMBL; AK029295; BAC03853.1; ALT\_SEQ.  
DR EMBL; AB058749; BAB47475.1; -  
DR Genew; HGNC:16091; ACAS2L.  
DR InterPro; IPR000873; AMP-bind.  
DR Pfam; PF00501; AMP-binding; 1.  
DR PROSITE; P500455; AMP\_BINDING; 1.  
KW Ligase; Mitochondrion; Transit  
FT TRANSIT 1 36 MITOCHONDRION (POTENTIAL).  
FT CHAIN 37 689 ACETYL-COENZYME A SYNTHETASE 2-LIKE.  
FT DOMAIN 45 53 POLY-ALA  
FT VARSPPLIC 446 447 Missing (in isoform 2).  
FT /FTIG=VSP\_007249.  
FT CONFLICT 277 277 V -> M (IN REF. 2; AAH39261).  
FT CONFLICT 488 488 V -> M (IN REF. 2; AAH44588).  
FT SEQUENCE 689 AA; 74856 MW; 66B84E39302AD08B CRC64;

```
Query Match          6.8%; Score 87; DB 1; Length 689;
Best Local Similarity 24.1%; Pred. No. 13;
Matches 33; Conservative 16; Mismatches 36; Indels 52; Gaps 7;

QY 9 FTGWPIVLVDGDVANGHKFVSVEGEGDATYKGLTKFKICTTGGKLPVPMPTLVTTLSYG 68
Db 473 FFGIVPVLMDKGSV-----VEGNSVSGALCIS-----QAWPGMARTI--- 510
QY 69 VQCFSPYDPMKQHDFFKSPMPGPGYVQERTIFFKDDGNKYTRA---EVKFEGLTVNRIE 125
Db 511 -----YGDHQRVDYAFKAYP-GY-----YFTGDGAYRTEGGYQITGRMDVDVI----- 553
QY 126 LKGDGDFKEDGNILGHKL 142
Db 554 -----NISGRL 560

RESULT 10
CP51_CANGA
ID_CP51_CANGA STANDARD; PRT; 533 AA.
AC PS0855; Q02312;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 51 (EC 1.14.13.70) (CYPLI) (P450-LIAl) (Sterol 14-
DE alpha-demethylase) (Lanosterol 14-alpha demethylase) (P450-14DM).
GN ERG11 OR CYP51.
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OC NCBI_TaxID=5478;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2001-L5;
RX MEDLINE=96161286; PubMed=8593007;
RA Geber A., Hitchcock C.A., Swartz J.E., Pullen F.S., Marsden K.E.,
RA Kwon-Chung K.J., Bennett J.E.;
RT "Deletion of the Candida glabrata ERG3 and ERG11 genes: effect on cell
RT viability, cell growth, sterol composition, and antifungal
RT susceptibility."
RL Antimicrob. Agents Chemother. 39:2708-2717(1995).
RN [2]
RP SEQUENCE OF 60-473 FROM N.A.
RC STRAIN=ATCC 2001;
RX MEDLINE=95081364; PubMed=7989540;
RA Burgerer-Kairuz P., Zuber J.P., Jaunin P., Buchman T.G., Bille J.,
RA Rossier M.;
RT "Rapid detection and identification of Candida albicans and
RT Torulopsis (Candida) glabrata in clinical specimens by
RT species-specific nested PCR amplification of a cytochrome P-450
RT lanosterol-alpha-demethylase (LIAL) gene fragment."
RL J. Clin. Microbiol. 32:1902-1907(1994).
CC -I- FUNCTION: Catalyzes C14-demethylation of lanosterol which is
CC critical for ergosterol biosynthesis. It transforms lanosterol
CC into 4,4'-dimethyl cholesta-8,14,24-triene-3-beta-ol (by
CC similarity).
CC -I- CATALYTIC ACTIVITY: Obtusifoliol + 3 O(2) + 3 NADPH = 4-alpha-
CC methyl-5-alpha-ergosta-8,14,24(28)-trien-3-beta-ol + formate + 3
CC NADP(+) + 3 H(2)O.
CC -I- SIMILARITY: Belongs to the cytochrome P450 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L40389; AAB02329.1; -
DR EMBL; S75389; AAB32679.1; -
DR
```

```
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Electron transport; Oxidoreductase; Monooxygenase; Membrane; Heme;
FT METAL 472 472 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT CONFLICT 64 64 I -> M (IN REF. 2).
FT CONFLICT 473 473 I -> T (IN REF. 2).
SQ SEQUENCE 533 AA; 61305 MW; A0506C17507B6EF7 CRC64;

Query Match          6.8%; Score 86.5; DB 1; Length 533;
Best Local Similarity 21.4%; Pred. No. 10;
Matches 44; Conservative 32; Mismatches 81; Indels 49; Gaps 8;

QY 25 GHKFSVS---GEGEGDATYKGLTKFKICTTGGKLPVPMPTLVTTLSYGVCFSRYPDH--M 79
Db 109 GHEFIFNAKLADVSAEAYSHL-----TTPVFGKGVYDCPNRLM 149
QY 80 KQHDFFKSAM-PEGYV-----QERTIFFKDDGNKYTRAEVKFEGLTVNRIELKGIDF 131
Db 150 EQKKFVKGALTKEAFVYVYVPLIAEEIYKYFRSNKFNENNSGIVDMVMSQPEM--TIF 207
QY 132 KEDGNILGHKLEYNHYNHVMADKQKGIKVNFKIRINIEDSGVOLADHYQONTPIGD 191
Db 208 TASRSLLGKEMRKDLDTDFAYLYSLDKGFTPINF-VFPNLPLEHYRKRDHAQAIS--- 263
QY 192 GPVLLPDNHYLSQTQSALSKDPNEKRD 217
Db 264 -----GTYNLSLIKREKND 278

RESULT 11
SYL_XYLFT
ID_SYL_XYLFT STANDARD; PRT; 879 AA.
AC Q87C65;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (Leurs).
GN LEUS OR PD1230.
OS Xylella fastidiosa (strain Temecual / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OC NCBI_TaxID=183190;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22421331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Teai S.M.,
RA Carrier H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA Coutinho L.B., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,
RA Marino C.L., Gaglioti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,
RA da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sassaki F.T., Sena J.A.D.,
RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,
RA Civerolo E.B., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA Kitajima J.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of Xylella
RT fastidiosa";
RL J. Bacteriol. 185:1018-1026(2003).
CC -I- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(leu).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; A8012557; AAO29080.1; ALT\_INIT.  
HAWAII; MF\_00049; -. 1.  
DR InterPro; IPR002302; Leu-TRNAsynt1a.  
DR InterPro; IPR002300; trNA-synt 1a.  
DR InterPro; IPR001412; trNA-synt 1.  
DR InterPro; IPR009008; ValRS IleRS\_edit.  
DR Pfam; PF00133; trNA-synt\_1; 1.  
DR PRINTS; PR00985; TRNASYNTHLEU.  
DR TIGRFAMs; TIGR00396; leuS bact; 1.  
DR PROSITE; PS00178; AA TRNA LIGASE I; 1.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KW Complete proteome.  
FT SITE 45 55 "HIGH" REGION.  
FT SITE 637 641 "RMSRS" REGION.  
FT BINDING 640 640 ATP (BY SIMILARITY).  
SQ SEQUENCE 879 AA; 99823 MW; 4C2BE01B8FDC497E CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 879;  
Best Local Similarity 22.2%; Pred. No. 18; Mismatches 69; Indels 57; Gaps 10;  
Matches 44; Conservative 28;  
50 TTGKLPVPWPTLVTLSYGVQCFSPYDPMKQHDFFKSAMPEGVYQERTIFPKDGNV-- 107  
Db 321 TNEQLPV-WVANFLMAYGTGAVMVGPHDQDEP--ANKYGLPTROVIALKEPKQDE 377  
QY 108 -----KTRAEVKEFGDTLVNRIELKIDFKEDGNILGHKLEYNNSHNVYI 153  
Db 378 STWEPDVRWDYADKTR---EFE---LNSAEPLDGLDYQAPVLAERFE----- 421  
QY 154 MADKQKNG-IKYNFKIRHNIEDGSVOLADHYQONTPI-----GDGPVLLPDN 199  
Db 422 ---RQGRQRRVYRLR-----DMGVSQRQYGCPIPIVYCTGAVVPENQLPVLPEN 474  
QY 200 -HYLSQTQSALSKDPNEKR 216  
Db 475 VAFSGTGSPIKTDPEWRK 492

RESULT 12  
VIT4 CAEEL STANDARD; PRT; 1603 AA.  
AC P18947; Q98BP3.  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Vitellogenin 4 precursor.  
GN VIT-4 OR F59D8.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 1-282 FROM N.A.  
RA Blumenthal T.; Spieth J.; Zucker E.;  
RL Submitted (NOV-1986) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 1-71 FROM N.A.  
RX MEDLINE=85269643; PubMed=4022780;  
RA Spieth J.; Denison K.; Kirtland S.; Cane J.; Blumenthal T.;  
RT "The C. elegans vitellogenin genes: short sequence repeats in the promoter regions and homology to the vertebrate genes."  
RL Nucleic Acids Res. 13:5283-5295(1985).  
CC -!- FUNCTION: Precursor of the egg-yolk proteins that are sources of

nutrients during embryonic development (Potential).  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- TISSUE SPECIFICITY: Synthesized in Caenorhabditis only by 32 cells building the intestine of adult hermaphroditic individuals; they are cotranslationally secreted into the body cavity and subsequently taken up by the gonad.  
CC -!- SIMILARITY: Contains 1 VWFD domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AC024137; AAK09074.1; -.  
EMBL; M11498; AAZ28163.1; -.  
EMBL; X02754; CAA26531.1; -.  
PIR; A43084; A43084.  
DR WormPep; F59D8.2; CE26817.  
DR InterPro; IPR001747; Lipid\_transprt\_N.  
DR InterPro; IPR001846; VWF\_D.  
DR Pfam; PF01347; Vitellogenin\_N; 1.  
DR SMART; SM00638; LPD\_N; 1.  
DR SMART; SM00216; VND; 1.  
KW Storage protein; Multigene family; Signal.  
FT SIGNAL 1 15 POTENTIAL.  
FT CHAIN 16 1603 VITELLOGENIN 4.  
FT DOMAIN 1308 1455 VWFD.  
FT CONFLICT 30 30 Y -> V (IN REF. 3).  
FT CONFLICT 169 169 L -> V (IN REF. 2).  
FT CONFLICT 183 187 EVAYT -> RSLRH (IN REF. 2).  
FT CONFLICT 275 275 T -> S (IN REF. 2).  
SQ SEQUENCE 1603 AA; 186307 MW; E3031703258C99BB CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 1603;  
Best Local Similarity 23.4%; Pred. No. 38;  
Matches 52; Conservative 32; Mismatches 69; Indels 69; Gaps 12;  
QY 1 MYVSGEELFTGVVPIVLVDGVDNGHKFSVSGEGSDATYGLTKLFICTTGKLPVWPMT 60  
Db 162 MESDKDSUFFNVHEKTMGDCV---AYTIVQEG-GKTIYTKSVNPKCITR-----PE 211  
QY 61 LVTTLISYGVQCFSPYDPMKQHDFFKSAMPEG-VYQERTIF---FKDDG----- 105  
Db 212 TAYGLRFGSEC-----KECEKEGFQVQPIVYTYTFKNEKLQSEVNSIYT 257  
QY 106 -----NYKTRAEVKEFGDTLVNRIELKIDFKEDGNILGHKLEYNNSHNVIMAD 156  
Db 258 LNVNGQEVVVKSETRAKVTVEESKINR-EIK-----KVSQPKKEIVYSMENEKLEIQ 308  
QY 157 KQKNG-----IKYNFKIRHNIEDGSVOLADHYQONTP 188  
Db 309 FYKQGDKAEPNFKAIETIEQKV-EQLEEIFRQIOBH-EQNTP 348

RESULT 13  
TRNA\_CAMJE STANDARD; PRT; 357 AA.  
AC Q9PP92;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE tRNA (Uracil-S)-methyltransferase (EC 2.1.1.35) (tRNA (M-5-U54)-methyltransferase) (RUMT).  
GN TRMA OR CJO831C.  
OS Campylobacter jejuni.  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
OC Campylobacteraceae; Campylobacter.  
OX NCBI\_TaxID=197;  
RN [1]  
RP SEQUENCE FROM N.A.





DE	Alkyl-dihydroxyacetonephosphate synthase, peroxisomal precursor
DE	(EC 2.5.1.26) (Alkyl-DHAP synthase) (Alkylglycerone-phosphate
DE	synthase).
GN	AGPS.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Liver;
RA	MEDLINE=98221159; PubMed=9187299;
RA	de Vet E.C.J.M., van den Broek B.T.E., van den Bosch H.;
RT	"Nucleotide sequence of human alkyl-dihydroxyacetonephosphate
RT	synthase cDNA reveals the presence of a peroxisomal targeting signal
RT	2.";
RL	Biochim. Biophys. Acta 1346:25-29(1997).
RN	[2]
RP	VARIANT RCDP3 HIS-419.
RX	MEDLINE=98221159; PubMed=9533082;
RA	de Vet E.C.J.M., Ijlst L., Oostheim W., Wanders R.J.A.,
RA	van den Bosch H.;
RT	"Alkyl-dihydroxyacetonephosphate synthase. Fate in peroxisome
RT	biogenesis disorders and identification of the point mutation
RT	underlying a single enzyme deficiency.";
RL	J. Biol. Chem. 273:10296-10301(1998).
CC	-!- CATALYTIC ACTIVITY: 1-acyl-glycerone 3-phosphate + a long-chain
CC	alcohol = 1-alkyl-glycerone 3-phosphate + a long-chain acid anion.
CC	-!- COFACTOR: FAD.
CC	-!- PATHWAY: Ether lipid biosynthesis.
CC	-!- SUBUNIT: May be part of an heterotrimeric complex composed of DAP-
CC	AT, ADAP-S and a modified form of DAP-AT.
CC	-!- SUBCELLULAR LOCATION: PEROXISOMAL. LOCALIZED TO THE INNER ASPECT
CC	OF THE PEROXISOMAL MEMBRANE.
CC	-!- DISEASE: Defects in AGPS are the cause of rhizomelic
CC	chondrodysplasia punctata, type 3 (RCDP3) [MIM:600121]. RCDP3 is
CC	an autosomal recessive disease characterized by rhizomelic
CC	shortening of the upper extremities, severe growth and mental
CC	retardation and cataract.
CC	-!- SIMILARITY: BELONGS TO THE FAD-BINDING OXIDOREDUCTASE/TRANSFERASE
CC	FAMILY 4.

Search completed: June 21, 2004, 15:55:25  
Job time : 7.55556 secs

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR	EMBL; Y09443; CAA70591.1; -
DR	Genew; HGNC:327; AGPS.
DR	MIM; 603051; -
DR	MIM; 600121; -
DR	GO; GO:0005777; C:peroxisome; TAS.
DR	GO; GO:0006629; P:lipid metabolism; TAS.
DR	InterPro; IPR004113; FAD-oxidase C.
DR	InterPro; IPR006094; Oxid FAD bind.
DR	Pfam; PF02513; FAD-oxidase_C; 1.
DR	Pfam; PF01565; FAD binding_4; 1.
KW	Lipid synthesis; Transferase; Flavoprotein; FAD; Peroxisome;
KW	Transit peptide; Rhizomelic chondrodysplasia punctata;
KW	Disease mutation.
FT	TRANSIT 1 58 MICROBODY (BY SIMILARITY).
FT	CHAIN 59 658 ALKYLDIHYDROXYACETONEPHOSPHATE SYNTHASE.
FT	DOMAIN 2 8 POLY-ALA.
FT	VARIANT 419 419 R -> H (in RCDP3).
FT	/FTid=VAR_005002.
SQ	SEQUENCE 658 AA; 72911 MW; 0E97AE86B513DF32 CRC64;

Query Match 6.7%; Score 85.5; DB 1; Length 658;  
Best Local Similarity 24.2%; Pred. NO. 16;  
Matches 45; Conservative 22; Mismatches 60; Indels 59; Gaps 10;



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:45:30 ; Search time 30.6667 Seconds  
(without alignments)  
2458.984 Million cell updates/sec

Title: US-09-887-784-222G

Perfect score: 1276

Sequence: 1 MVSKGELFTGVVPIVLVD.....VLGFTVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.25.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phage.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriaph.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1235	96.8	238	2	Q8GHE2
2	1232	96.6	238	5	Q93125
3	1230	96.4	238	2	Q8GHE4
4	1229	96.3	238	2	Q8GHE3
5	1197	93.8	238	5	Q17105
6	1182	92.6	238	5	Q17106
7	1077	84.4	238	5	Q8WTC6
8	1073	84.1	238	5	Q8WTP95
9	1069	83.1	238	5	Q8WTC4
10	1067	83.6	238	5	Q8WTD0
11	1066	83.5	238	5	Q8WTC8
12	1066	83.5	238	5	Q8WTC9
13	1064	83.4	238	5	Q8WTC7
14	1062	83.2	238	5	Q8WTC5
15	252.5	19.8	225	5	Q95UA7
16	252.5	19.8	225	5	Q720W5

17	247	19.4	225	5	Q963F5	Q963F5	montastraea
18	244.5	19.2	236	5	Q8T6U0	Q8T6U0	dendronepht
19	242.5	19.0	225	5	Q720W9	Q720W9	montastraea
20	240	18.8	225	5	Q816J8	Q816J8	trachyphyll
21	238.5	18.7	266	5	Q9U6Y3	Q9U6Y3	clavularia
22	233	18.3	225	5	Q720W4	Q720W4	montastraea
23	232	18.2	224	5	Q8MU48	Q8MU48	montastraea
24	232	18.2	225	5	Q8T5F1	Q8T5F1	montastraea
25	214	16.8	227	5	Q720W6	Q720W6	montastraea
26	214	16.8	234	5	Q720W7	Q720W7	montastraea
27	212.5	16.7	259	5	Q8MMA2	Q8MMA2	agaricia fr
28	210	16.5	239	5	Q8MMA1	Q8MMA1	agaricia ag
29	209	16.4	234	5	Q8T5F2	Q8T5F2	montastraea
30	209	16.4	234	5	Q8MU47	Q8MU47	montastraea
31	208.5	16.3	229	5	Q9U6Y6	Q9U6Y6	anemonia ma
32	206.5	16.2	232	5	Q9GZ28	Q9GZ28	anemonia su
33	206	16.1	227	5	Q962P9	Q962P9	montastraea
34	206	16.1	227	5	Q720W8	Q720W8	montastraea
35	205.5	16.1	232	5	Q9GP15	Q9GP15	anemonia su
36	204.5	16.0	214	5	Q86LV7	Q86LV7	meandrina m
37	204	16.0	221	5	Q95P04	Q95P04	goniopora t
38	203.5	15.9	214	5	Q86LV8	Q86LV8	meandrina m
39	203.5	15.9	238	5	Q9BLV9	Q9BLV9	renilla mue
40	202	15.8	227	5	Q95VT0	Q95VT0	montastraea
41	201.5	15.8	225	5	Q9U6Y8	Q9U6Y8	discosoma s
42	198.5	15.6	222	5	Q72168	Q72168	cerianthus
43	198.5	15.6	225	5	Q8T6T9	Q8T6T9	radianthus
44	198	15.5	235	5	Q8T5F0	Q8T5F0	scolymia cu
45	197.5	15.5	232	5	Q9U6Y7	Q9U6Y7	discosoma s

ALIGNMENTS

RESULT 1

Q8GHE2	PRELIMINARY;	PRT;	238 AA.
ID	Q8GHE2		
AC	Q8GHE2;		
DC	01-MAR-2003 (Tremblrel. 23, Created)		
DT	01-MAR-2003 (Tremblrel. 23, Last sequence update)		
DT	01-OCT-2003 (Tremblrel. 25, Last annotation update)		
DE	Green fluorescence protein.		
GN	2289GFP.		
OS	Azotobacter vinelandii.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;		
OC	Pseudomonadaceae; Azotobacter.		
OX	NCBI_TaxID=354;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=DSM2289;		
RA	Koranyi P., Berenyi M., Burg K.;		
RT	"Occurrence of green fluorescence protein in diazotrophic bacteria		
RT	Azomonas and Azotobacter."		
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF324408; AAN86140.1; -		
DR	GO; GO:0006091; P.energy pathways; IEA.		
DR	InterPro; IPR009017; GFP_like.		
DR	InterPro; IPR000786; Green_Fl_protein.		
DR	Pfam; PF01353; GFP; 1.		
DR	PRINTS; PR01229; GFPLORESCENT.		
DR	ProDom; PD013756; Green_fl_protein; 1.		
SQ	SEQUENCE 238 AA; 26914 MW; F84840F1F9064018 CRC64;		

Query Match 96.8%; Score 1235; DB 2; Length 238;  
Best Local Similarity 97.5%; Pred. No. 1.7e-94;  
Matches 232; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy	2	VSKGELFTGVVPIVLVDGVNGHKFSVSGEGDATYVKLTKEICTTGKLPVWPPTL	61
Db	1	MSKGELFTGVVPIVLVDGVNGHKFSVSGEGDATYVKLTKEICTTGKLPVWPPTL	60
Qy	62	VTTLSVGVCFSRYPDHMKOHDFFKSAMPEGVVQERTIFFKDDGNKYKRAEVKFGDTLV	121

```
Db      61 VTTFSVGVCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKODGNKYKTRAEVKFEGDTLV 120
QY      122 NRIELKIGDFKEDGNILGHKLEYNNSHNYYIMADKQKNGIKVNFIRHNIEDGSVOLAD 181
Db      121 NRIELKIGDFKEDGNILGHKLEYNNSHNYYIMADKQKNGIKVNFIRHNIEDGSVOLAD 180
QY      182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGGFVTAAGITLGMDELYK 239
Db      181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGGFVTAAGITLGMDELYK 238

RESULT 2
Q93125  PRELIMINARY; PRT; 238 AA.
AC      Q93125;
DT      01-FEB-1997 (TrEMBLrel. 02, Created)
DT      01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Green fluorescent protein mutant 3.
GN      GFP.
OS      Aequorea victoria (Jellyfish).
OC      Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC      Aequoreidae; Aequorea.
OX      NCBI_TaxID=6100;
RN      [1]_TaxID=6100;
RP      SEQUENCE FROM N.A.
RX      MEDLINE=96305137; PubMed=8707053;
RA      Cormack B.P., Valdivia R.H., Falkow S.;
RT      "FACS-optimized mutants of the green fluorescent protein (GFP).";
RL      Gene 173:33-38(1996).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      Cormack B.P., Bertram G., Egerton M., Gow N.A.R., Falkow S.,
RA      Brown A.J.P.;
RT      "Yeast Enhanced Green Fluorescent Protein (yEGFP): a reporter of gene
RT      expression in Candida albicans.";
RL      Microbiology 0:0-0(1996).
DR      EMBL; U73901; AAB18957.1; -.
DR      HSSP; P42232; 1BFP.
DR      GO; GO:0006091; P:energy pathways; IEA.
DR      InterPro; IPR009017; GFP_like.
DR      Pfam; PF01353; GFP; 1.
DR      PRINTS; PR01229; GFPLORESCENT.
DR      ProDom; PD013756; Green_fl_protein; 1.
DR      SEQUENCE 238 AA; 26840 MW; A28622809A9DERA60 CRC64;

Query Match          96.6%; Score 1232; DB 5; Length 238;
Best Local Similarity 97.1%; Pred. No. 3e-94;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY      2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKLFICTTGKLPVWPPTL 61
Db      1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKLFICTTGKLPVWPPTL 60
QY      62 VTTLSVGVCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKODGNKYKTRAEVKFEGDTLV 121
Db      61 VTTFSVGVCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKODGNKYKTRAEVKFEGDTLV 120
QY      122 NRIELKIGDFKEDGNILGHKLEYNNSHNYYIMADKQKNGIKVNFIRHNIEDGSVOLAD 181
Db      121 NRIELKIGDFKEDGNILGHKLEYNNSHNYYIMADKQKNGIKVNFIRHNIEDGSVOLAD 180
QY      182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGGFVTAAGITLGMDELYK 239
Db      181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGGFVTAAGITLGMDELYK 238

RESULT 3
Q8GHE4  PRELIMINARY; PRT; 238 AA.
AC      Q8GHE4;
DT      01-MAR-2003 (TrEMBLrel. 23, Created)
```

```
DT      01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Green fluorescence protein.
GN      375GFP.
OS      Azomonas agilis.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC      Pseudomonadaceae; Azomonas.
OX      NCBI_TaxID=116849;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Koranyi P., Berenyi M., Burg K.;
RT      "Occurrence of green fluorescence protein in diazotrophic bacteria
RT      Azomonas and Azotobacter.";
RL      Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF324405; AAN86137.1; -.
DR      GO; GO:0006091; P:energy pathways; IEA.
DR      InterPro; IPR009017; GFP_like.
DR      InterPro; IPR000786; Green_fl_protein.
DR      Pfam; PF01353; GFP; 1.
DR      PRINTS; PR01229; GFPLORESCENT.
DR      ProDom; PD013756; Green_fl_protein; 1.
DR      SEQUENCE 238 AA; 26902 MW; 15FE9B9C5B4F6B89 CRC64;

Query Match          96.4%; Score 1230; DB 2; Length 238;
Best Local Similarity 97.1%; Pred. No. 4.4e-94;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY      2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKLFICTTGKLPVWPPTL 61
Db      1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKLFICTTGKLPVWPPTL 60
QY      62 VTTLSVGVCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKODGNKYKTRAEVKFEGDTLV 121
Db      61 VTTFSVGVCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKODGNKYKTRAEVKFEGDTLV 120
QY      122 NRIELKIGDFKEDGNILGHKLEYNNSHNYYIMADKQKNGIKVNFIRHNIEDGSVOLAD 181
Db      121 NRIELKIGDFKEDGNILGHKLEYNNSHNYYIMADKQKNGIKVNFIRHNIEDGSVOLAD 180
QY      182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGGFVTAAGITLGMDELYK 239
Db      181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGGFVTAAGITLGMDELYK 238

RESULT 4
Q8GHE3  PRELIMINARY; PRT; 238 AA.
AC      Q8GHE3;
DT      01-MAR-2003 (TrEMBLrel. 23, Created)
DT      01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Green fluorescence protein.
GN      85GFP.
OS      Azotobacter vinelandii.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC      Pseudomonadaceae; Azotobacter.
OX      NCBI_TaxID=354;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Koranyi P., Berenyi M., Burg K.;
RT      "Occurrence of green fluorescence protein in diazotrophic bacteria
RT      Azomonas and Azotobacter.";
RL      Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF324406; AAN86138.1; -.
DR      GO; GO:0006091; P:energy pathways; IEA.
DR      InterPro; IPR009017; GFP_like.
DR      InterPro; IPR000786; Green_fl_protein.
DR      Pfam; PF01353; GFP; 1.
DR      PRINTS; PR01229; GFPLORESCENT.
DR      ProDom; PD013756; Green_fl_protein; 1.
DR      SEQUENCE 238 AA; 26887 MW; E0E1616BD2AF6188 CRC64;

Query Match          96.3%; Score 1229; DB 2; Length 238;
```

```
Best Local Similarity 97.1%; Pred. No. 5.4e-94;
Matches 231; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYKGLTLKFICTTGKLPVWPPTL 60

QY 62 VTTLSTGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQKNGIKVNFIRHNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQKNGIKVNFIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLVLEFVTAAGITHGMDELYK 238

RESULT 5
Q17105 PRELIMINARY; PRT; 238 AA.
ID Q17105
AC Q17105
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83959; CAA58789.1; -.
DR PIR; J05692; J01514.
DR HSSP; P42212; 1BFP.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP_like.
DR Pfam; PF01353; GFP. 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748E44 CRC64;

Query Match 93.8%; Score 1197; DB 5; Length 238;
Best Local Similarity 93.7%; Pred. No. 2.4e-91;
Matches 223; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYKGLTLKFICTTGKLPVWPPTL 60

QY 62 VTTLSTGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQKNGIKVNFIRHNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQKNGIKVNFIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLVLEFVTAAGITHGMDELYK 238

RESULT 6
Q17106 PRELIMINARY; PRT; 238 AA.
ID Q17106
```

```
AC Q17106;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83960; CAA58790.1; -.
DR PIR; J05692; J01514.
DR HSSP; P42212; 1BFP.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP_like.
DR Pfam; PF01353; GFP. 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; BD4648262D8EABD4 CRC64;

Query Match 92.6%; Score 1182; DB 5; Length 238;
Best Local Similarity 92.9%; Pred. No. 4.2e-90;
Matches 221; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYKGLTLKFICTTGKLPVWPPTL 60

QY 62 VTTLSTGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQKNGIKVNFIRHNIEDGSVOLAD 181
DB 121 NRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQKNGIKVNFIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLVGGFVTAAGITLGMDELYK 239
DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDHMLVLEFVTAAGITHGMDELYK 238

RESULT 7
Q8WTC6 PRELIMINARY; PRT; 238 AA.
ID Q8WTC6
AC Q8WTC6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435431; AAL33916.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP_like.
DR Pfam; PF01353; GFP. 1.
```

```
DR PRINTS; PR01229; GFP.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27015 MW; 6B8FD75E88926903 CRC64;

Query Match      84.4%; Score 1077; DB 5; Length 238;
Best Local Similarity 81.9%; Pred. No. 2.2e-81;
Matches 195; Conservative 20; Mismatches 23; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 60

QY 62 VTTLSTGVQCFSPYDPHMKOHDFKSAFPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB 61 VTTLSTGVQCFARYPEHMKNDFFKSAFPEGYIQERTIFFQDDGKYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFIRHNIEDGSVOLAD 181
DB 121 NRIELKGMDFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVGGFVTAAGITLGMDELYK 239
DB 181 HYQTNVPLGDGPVLLPDNHYLSTQTSALSADPNKRDHMLVGGFVTAAGITLGMDELYK 238

RESULT 8
Q8WP95 PRELIMINARY; PRT; 238 AA.
AC Q8WP95;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFPXM.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL33918.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match      83.8%; Score 1069; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 1e-80;
Matches 196; Conservative 17; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 60

QY 62 VTTLSTGVQCFSPYDPHMKOHDFKSAFPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB 61 VTTLSTGVQCFARYPEHMKNDFFKSAFPEGYIQERTIFFQDDGKYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFIRHNIEDGSVOLAD 181
DB 121 NRIELKGMDFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVGGFVTAAGITLGMDELYK 239
DB 181 HYQTNVPLGDGPVLLPDNHYLSTQTSALSADPNKRDHMLVGGFVTAAGITLGMDELYK 238

RESULT 10
Q8WTD0 PRELIMINARY; PRT; 238 AA.
AC Q8WTD0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea macrodactyla."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY013824; AAK02062.1; -.
DR EMBL; AY013821; AAK02059.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27049 MW; 8185D0E5E529012B CRC64;

Query Match      84.1%; Score 1073; DB 5; Length 238;
Best Local Similarity 81.5%; Pred. No. 4.6e-81;
Matches 194; Conservative 20; Mismatches 24; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 61
DB 1 MSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 60

QY 62 VTTLSTGVQCFSPYDPHMKOHDFKSAFPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
DB 61 VTTLSTGVQCFARYPEHMKNDFFKSAFPEGYIQERTIFFQDDGKYKTRAEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFIRHNIEDGSVOLAD 181
DB 121 NRIELKGMDFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLVGGFVTAAGITLGMDELYK 239
DB 181 HYQTNVPLGDGPVLLPDNHYLSTQTSALSADPNKRDHMLVGGFVTAAGITLGMDELYK 238
```

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF435427; AAL33912.1; --  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR009017; GFP\_like.  
DR InterPro; IPR000786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFLUORESCENT.  
DR ProDom; PD013756; Green fl protein; 1.  
SQ SEQUENCE 238 AA; 26997 MW; 5F80A192173CB84D CRC64;

Query Match 83.5%; Score 1067; DB 5; Length 238;

Best Local Similarity 81.1%; Pred. No. 1.8e-80;  
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVDGNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPPTL 61  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
1 MSKGEELFTGIVPVLIELDGDVGHKFSVRGEGDADYGLKLEIKFICTTGKLPVPWPPTL 60  
QY 62 VTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
61 VTTLGYGIQCFARYPEHMKMNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120  
QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
121 NRIELKGMDFKEDGNILGHKLEYNHNHYIMPDKANNGLVNFKIRHNIEGGVQOLAD 180  
QY 182 HYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLVGGFVTAAGITLGMDELYK 239  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
181 HYQTNVPLGDGVPVLPINHYLSQTALSKDRNETRDHMLVLEFFSACGTHGMDELYK 238

## RESULT 11

Q8WTC8 PRELIMINARY; PRT; 238 AA.

AC Q8WTC8; 01-MAR-2002 (TremBLrel. 20, Created)  
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Green fluorescent protein.  
GN GFP.

OS Aequorea macrodactyla.  
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroida; Leptomedusae;  
OC Aequoreidae; Aequorea.

OX NCBI\_TaxID=147615;

RN [1]\_TaxID=147615;

RP SEQUENCE FROM N.A.

RC STRAIN=GFPxm163;

RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,

Li S.J., Xia N.S.;

RT "Colorful mutants of green fluorescent protein from Aequorea

macrodactyla.";

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF435429; AAL33914.1; --

DR GO; GO:0006091; P:energy pathways; IEA.

DR InterPro; IPR009017; GFP\_like.

DR InterPro; IPR000786; Green\_fl\_protein.

DR Pfam; PF01353; GFP; 1.

DR PRINTS; PR01229; GFLUORESCENT.

DR ProDom; PD013756; Green fl protein; 1.

SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;

Query Match

Best Local Similarity 83.5%; Score 1066; DB 5; Length 238;

Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVDGNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPPTL 61  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
1 MSKGEELFTGIVPVLIELDGDVGHKFSVRGEGDADYGLKLEIKFICTTGKLPVPWPPTL 60  
QY 62 VTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
61 VTTLGYGIQCFARYPEHMKMNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
121 NRIELKGMDFKEDGNILGHKLEYNHNHYIMPDKANNGLVNFKIRHNIEGGVQOLAD 180  
QY 182 HYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLVGGFVTAAGITLGMDELYK 239  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
181 HYQTNVPLGDGVPVLPINHYLSQTALSKDRNETRDHMLVLEFFSACGTHGMDELYK 238

## RESULT 12

Q8WTC9 PRELIMINARY; PRT; 238 AA.

AC Q8WTC9; 01-MAR-2002 (TremBLrel. 20, Created)  
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Green fluorescent protein.  
GN GFP.

OS Aequorea macrodactyla.  
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroida; Leptomedusae;  
OC Aequoreidae; Aequorea.

OX NCBI\_TaxID=147615;

RN [1]\_TaxID=147615;

RP SEQUENCE FROM N.A.

RC STRAIN=GFPxm162;

RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,

Li S.J., Xia N.S.;

RT "Colorful mutants of green fluorescent protein from Aequorea

macrodactyla.";

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF435428; AAL33913.1; --

DR GO; GO:0006091; P:energy pathways; IEA.

DR InterPro; IPR009017; GFP\_like.

DR InterPro; IPR000786; Green\_fl\_protein.

DR Pfam; PF01353; GFP; 1.

DR PRINTS; PR01229; GFLUORESCENT.

DR ProDom; PD013756; Green fl protein; 1.

SQ SEQUENCE 238 AA; 27031 MW; 5F80A19C19DC584D CRC64;

Query Match

Best Local Similarity 83.5%; Score 1066; DB 5; Length 238;

Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVDGNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWPPTL 61  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
1 MSKGEELFTGIVPVLIELDGDVGHKFSVRGEGDADYGLKLEIKFICTTGKLPVPWPPTL 60  
QY 62 VTLLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
61 VTTLGYGIQCFARYPEHMKMNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120  
QY 122 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
121 NRIELKGMDFKEDGNILGHKLEYNHNHYIMPDKANNGLVNFKIRHNIEGGVQOLAD 180  
QY 182 HYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLVGGFVTAAGITLGMDELYK 239  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
181 HYQTNVPLGDGVPVLPINHYLSQTALSKDRNETRDHMLVLEFFSACGTHGMDELYK 238

## RESULT 13

Q8WTC7 PRELIMINARY; PRT; 238 AA.

AC Q8WTC7; 01-MAR-2002 (TremBLrel. 20, Created)  
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Green fluorescent protein.  
GN GFP.

OS Aequorea macrodactyla.  
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroida; Leptomedusae;  
OC Aequoreidae; Aequorea.

OX NCBI\_TaxID=147615;

Db 1 MSKGEELFTGVVPILVELDGDVGHGKFTSVRGESEGDADYGKLEIKFICTTGKLPVPWPTL 60  
Qy 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKKEGDTLV 121  
Db 61 VTTILGYLGFARYPEHMKQNDFFKSAMPEGYIQRITFFQDDGKYKTRGEVKEGDTLV 120  
Qy 122 NRLELKGIDPKEDGNILGHKLEYNVSHNVIMADKQKNGIKVNFKIRHNIEDGSVQLAD 181  
Db 121 NRLELKGMDFKEDGNILGHKLEYNFNSHNVIMPDKANNGLKVNFKIRHNIEGGVQLAD 180  
Qy 182 HYQONTPTIGDGPVLLPDNHYLSTQSALSADKDNKRDHMLVGGFVTAAGITLGMDELYK 239  
Db 181 HYQTNVPLGDGPVLLIPINHYLSQTAISKDRNTRDHMVLEFFSACGHTHGMDELYK 238

RESULT 15  
Q95UA7 PRELIMINARY; PRT; 225 AA.  
ID Q95UA7 AC Q95UA7; 01-DEC-2001 (TremBLrel. 19, Created)  
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Cytan fluorescent protein (Fragment).  
OS Montastraea cavernosa (great star coral).  
OC Eukaryota; Metazoa; Chnidaria; Anthozoa; Zoantharia; Scleractinia;  
OC Faviina; Faviidae; Montastraea.  
OX NCBI\_TaxID=63558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Falkowski P.G., Sun Y.;  
RT "Montastraea cavernosa fluorescent protein.";  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY056460; AAL17905.1; -  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR009017; GFP\_like.  
DR DR InterPro; IPR000786; Green\_fl\_protein.  
DR PFam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFLUORESCENT.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
FT NON TER 225 225  
SQ SEQUENCE 225 AA; 25775 MW; 52DE2F716D083524 CRC64;

Qy 12 VVPILVELDGDVNGHKFSVSGEGEDATYGLTKLF-ICTTGKLPVPWPTLVTTLSYGVC 70  
Db 7 VMKIKURMDGIVNGHGFMTIGEKGKFFGTHIILKVKGGPLPPAYDILTATFAQYGNR 66  
Qy 71 CFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKKEGDTLVNLELKGID 130  
Db 67 VFTKYPKDI-P-DYFKQSPPEGYSWERSMTFFDQGVCTVTSIDKLEGGDCFFYRFGVN 124  
Qy 131 FKDDGNILGHK-LEYNVNHNHVMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPI 189  
Db 125 FPSGGVPMQKTLKWEPTENMYV-----RDGVLLGDVSKTLLEGD----KHKRCNFRS 175  
Qy 190 GDGP---VLLPDNHYLSTQ-SALSAD 211  
Db 176 TYGAKKGVLPEYHFVDHRIEILSHD 201

Search completed: June 21, 2004, 16:00:13  
Job time : 31.7778 secs

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-GFPxm191uv;  
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,  
RA Li S.J., Xia N.S.;  
RT "Colorful mutants of green fluorescent protein from Aequorea  
RT macrodactyla.";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF435430; AAL33915.1; -  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR009017; GFP\_like.  
DR DR InterPro; IPR000786; Green\_fl\_protein.  
DR PFam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFLUORESCENT.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
SQ SEQUENCE 238 AA; 27002 MW; BD5BA2982264C018 CRC64;

Query Match 83.4%; Score 1064; DB 5; Length 238;  
Best Local Similarity 81.1%; Pred. No. 2.6e-80;  
Matches 193; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

Qy 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLFICTTGKLPVPWPTL 61  
Db 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDADYGLTKLFICTTGKLPVPWPTL 60  
Qy 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKKEGDTLV 121  
Db 61 VTTILGYLGFARYPEHMKQNDFFKSAMPEGYIQRITFFQDDGKYKTRGEVKEGDTLV 120  
Qy 122 NRLELKGIDPKEDGNILGHKLEYNVSHNVIMADKQKNGIKVNFKIRHNIEDGSVQLAD 181  
Db 121 NRLELKGMDFKEDGNILGHKLEYNFNSHNVIMPDKANNGLKVNFKIRHNIEGGVQLAD 180  
Qy 182 HYQONTPTIGDGPVLLPDNHYLSTQSALSADKDNKRDHMLVGGFVTAAGITLGMDELYK 239  
Db 181 HYQTNVPLGDGPVLLIPINHYLSQTAISKDRNTRDHMVLEFFSACGHTHGMDELYK 238

RESULT 14  
Q8WTC5 PRELIMINARY; PRT; 238 AA.  
ID Q8WTC5 AC Q8WTC5; 01-MAR-2002 (TremBLrel. 20, Created)  
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Orange fluorescent protein.  
GN GFP.  
OS Aequorea macrodactyla.  
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydrozoa; Leptomedusae;  
OC Aequoreidae; Aequorea.  
OX NCBI\_TaxID=147615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OPFxm;  
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,  
RA Li S.J., Xia N.S.;  
RT "Colorful mutants of green fluorescent protein from Aequorea  
RT macrodactyla.";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF435432; AAL33917.1; -  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR009017; GFP\_like.  
DR DR InterPro; IPR000786; Green\_fl\_protein.  
DR PFam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFLUORESCENT.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
SQ SEQUENCE 238 AA; 27018 MW; 75521EA5534E573A CRC64;

Query Match 83.2%; Score 1062; DB 5; Length 238;  
Best Local Similarity 81.9%; Pred. No. 3.8e-80;  
Matches 195; Conservative 17; Mismatches 26; Indels 0; Gaps 0;

Qy 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLFICTTGKLPVPWPTL 61



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:41:49 ; Search time 47.1111 Seconds  
(without alignments)  
1433.395 Million cell updates/sec

Title: US-09-887-784-222I

Perfect score: 1274

Sequence: 1 MYSKGEELFTGVVILVELD.....VLIGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: Genesecp1980s:\*
- 2: Genesecp1990s:\*
- 3: Genesecp2000s:\*
- 4: Genesecp2001s:\*
- 5: Genesecp2002s:\*
- 6: Genesecp2003as:\*
- 7: Genesecp2003bs:\*
- 8: Genesecp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1272	99.8	239	5	Aae17518 Enhanced
2	1272	99.8	363	6	Abr40352 Human ami
3	1272	99.8	893	4	Ag65781 Amino aci
4	1272	99.8	1132	4	Ag65782 Amino aci
5	1284	99.2	239	5	Aae17517 Enhanced
6	1261	99.0	239	3	Aab22882 Enhanced
7	1261	99.0	239	3	Aay54349 Amino aci
8	1261	99.0	239	3	Aay79584 EGFP sign
9	1261	99.0	239	4	Aab50804 Jellyfish
10	1261	99.0	239	4	Aab59900 A. victor
11	1261	99.0	239	4	Aab31171 Amino aci
12	1261	99.0	239	5	Ag66198 A. victor
13	1261	99.0	239	5	Ag69444 Protease
14	1261	99.0	239	5	Aae14599 Aequorea
15	1261	99.0	239	6	Aae14958 Aequorea
16	1261	99.0	239	6	Ag79829 Green flu
17	1261	99.0	239	6	Abr3616 Green flu
18	1261	99.0	239	6	Ada38074 Aequorea
19	1261	99.0	239	7	Abu63204 Aequorea
20	1261	99.0	239	7	Adc18358 EGFP (enh
21	1261	99.0	239	7	Abw00914 Aequorea
22	1261	99.0	239	7	Ade28570 Enhanced
23	1261	99.0	246	7	Abm79011 Enhanced
24	1261	99.0	248	5	Ag68319 Jellyfish
25	1261	99.0	259	5	Aau99804 Biomembra

ALIGNMENTS

RESULT 1

AAE17518  
ID AAE17518 standard; protein; 239 AA.

XX AAE17518;

DT XX  
22-APR-2002 (first entry)

DE Enhanced F64L-E222G jellyfish green fluorescent protein mutant.

KW Jellyfish; green fluorescent protein; GFP; protein redistribution;  
KW cellular function; genetic reporter; mutant; Stoke's shift; muten.

XX Aequorea victoria.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 65

FT /note= "Wild type Phe substituted with Leu; This  
corresponds to position 64 in the wild type protein"

FT Misc-difference 223  
FT /note= "Wild type Glu substituted with Gly; This  
corresponds to position 222 in the wild type protein"

FT WO200198338-A2.

PD 27-DEC-2001.

XX 18-JUN-2001; 2001WO-BP006848.

XX 19-JUN-2000; 2000DK-00000953.

PR 20-JUN-2000; 2000US-0213681P.

PR 10-MAY-2001; 2001DK-00000739.

PR 10-MAY-2001; 2001US-0290170P.

XX (BIOI-) BIOIMAGE AS.

XX Bjorn SP, Pagliaro L, Thastrup O;

XX WPI; 2002-098224/13.

DR N-PSDB; AAD28163.

XX Novel fluorescent protein in in vitro assay for measuring protein kinase

PT activity or dephosphorylation activity, or for measuring protein

PT redistribution, has a green fluorescent protein with F64L and E222G

PT mutation.

XX Claim 9; Page 37; 41pp; English.

XX The invention relates to a fluorescent protein derived from green  
 CC fluorescent protein (GFP) or its analogue. The GFP containing mutations  
 CC at F64L and E222G has a bigger compared to other GFP's making it very  
 CC suitable for high throughput screening due to better resolution. The  
 CC fluorescent protein is useful in invitro assays for measuring protein  
 CC kinase activity or dephosphorylation activity, or for measuring protein  
 CC redistribution. The fluorescent protein is useful in studying cellular  
 CC functions in living cells; as protein tags in transgenic animals, living  
 CC and fixed cells; organelle tags, secretion marker and genetic reporter.  
 CC The fluorescent protein is also useful as a cell or organelle integrity  
 CC marker, a marker for changes in cell morphology, as transfection marker,  
 CC and as a marker to be used in combination with fluorescence activated  
 CC cell sorting (FACS). The novel proteins can also be used as reporters to  
 CC monitor live or dead biomass of organisms, such as fungi. The fluorescent  
 CC protein is also useful as markers in transcriptional and translational  
 CC fusions for performing transposon vector mutagenesis and as a reporter  
 CC for bacterial detection. Transposons encoding the fluorescent protein are  
 CC useful for screening promoters and for tagging plasmids and chromosomes.  
 CC The fluorescent protein engineered into the genome of a phage is useful  
 CC for designing diagnostic tool. The present sequence is a DNA encoding  
 CC enhanced F64L-E222G jellyfish green fluorescent protein (GFP) mutant  
 XX Sequence 239 AA;

Query Match 99.8%; Score 1272; DB 5; Length 239;  
 Best Local Similarity 99.6%; Pred. No. 1.5e-122;  
 Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60  
 DB 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60  
 QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNHVIIMADKQKNGIKVNFKIRHNIEDGSVOLA 180  
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNHVIIMADKQKNGIKVNFKIRHNIEDGSVOLA 180  
 QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLTGFTVTAAGITLGMDELYK 239  
 DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLTGFTVTAAGITLGMDELYK 239

RESULT 2  
 ABR40352  
 ID ABR40352 standard; protein; 363 AA.  
 AC ABR40352;  
 XX 08-JUL-2003 (first entry)  
 DT Human amino acid sequence SEQ ID NO: 6.  
 XX Human; heterologous conjugate; intracellular protein.  
 XX Homo sapiens.  
 OS Aequorea victoria.  
 XX WO2002:029827-A2.  
 PN 10-APR-2003.  
 PD 01-OCT-2002; 2002WO-DK000651.  
 FF 01-OCT-2001; 2001DK-00001433.  
 PR 11-OCT-2001; 2001US-0328896P.  
 XX (BIOL-) BIOMAGE AS.  
 PA Terry BR, Nielsen SJ;  
 XX

XX WPI; 2003-430211/40.  
 DR N-PSDB; ACC72604.  
 XX Novel cell for identifying modulators of protein interaction, contains a  
 CC first conjugate comprising anchor protein, second conjugate having type B  
 CC interactor protein and third conjugate with detectable group.  
 PT Disclosure; Page 112-113; 118pp; English.  
 XX The invention relates to a novel cell, comprising three heterologous  
 CC conjugates (HC), a first HC (HC1) comprising an anchor protein that  
 CC specifically binds to an internal structure within the cell conjugated to  
 CC an interactor protein (IP) of type A, a second HC (HC2) comprising IP of  
 CC type B conjugated to a first protein of interest, and a third HC (HC3)  
 CC comprising a second protein of interest conjugated to detectable group.  
 CC The cell is useful for detecting if a compound disrupts or induces the  
 CC interaction between two intracellular proteins. The cell is also useful  
 CC for screening compounds that modulate the interaction between two  
 CC intracellular proteins. The present sequence is used in the  
 CC exemplification of the invention  
 XX Sequence 363 AA;

Query Match 99.8%; Score 1272; DB 6; Length 363;  
 Best Local Similarity 99.6%; Pred. No. 2.8e-122;  
 Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60  
 DB 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60  
 QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNHVIIMADKQKNGIKVNFKIRHNIEDGSVOLA 180  
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNHVIIMADKQKNGIKVNFKIRHNIEDGSVOLA 180  
 QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLTGFTVTAAGITLGMDELYK 239  
 DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLTGFTVTAAGITLGMDELYK 239

RESULT 3  
 AAG65781  
 ID AAG65781 standard; protein; 893 AA.  
 XX AAG65781;  
 AC AAG65781;  
 XX 07-JAN-2002 (first entry)  
 DT Amino acid sequence of HSPDE4A1-E222G fusion protein.  
 XX PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;  
 KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;  
 KW fusion protein.  
 XX Homo sapiens.  
 OS Aequorea victoria.  
 XX WO2001:79526-A2.  
 PN 25-OCT-2001.  
 PD 11-APR-2001; 2001WO-DK000264.  
 FF 17-APR-2000; 2000DK-00000651.  
 PR 29-MAY-2000; 2000DK-00000849.  
 XX (BIOI-) BIOIMAGE AS.  
 PA  
 XX

```
PI Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
PI Praestegaard M;
XX
XX WPI: 2001-611727/70.
DR N-PSDB; AA166852.
XX
XX Determining if a compound is a dislocator of PDE4 for identifying
PT compounds for treating CNS and inflammatory disease comprises identifying
PT compounds which remove PDE4 spots.
XX
XX Example 1; Page 156-160; 160pp; English.
XX
XX The invention relates to determining, if a compound, is a dislocator of
CC PDE4. The method comprises testing if the compound removes PDE4 spots,
CC which may optionally be induced by a Rolipram-like reference compound,
CC and testing if it inhibits the catalytic activity of the PDE4, where the
CC compound is a dislocator of PDE4, if it removes PDE spots and if it does
CC not inhibit the catalytic activity of PDE4. The method is useful for
CC identifying compounds useful for the treatment of diseases of the central
CC nervous system such as depression and for the treatment of inflammatory
CC disease such as joint inflammation, Crohn's disease, inflammatory bowel
CC disease, respiratory diseases, chronic obstructive pulmonary disease
CC (COPD), including asthma, chronic bronchitis, pulmonary emphysema,
CC endotoxemic shock, toxic shock syndrome, systemic lupus erythematosus,
CC psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
CC infection. The use of a reagent that can mimic or reverse the effect of
CC the compound with affinity for the catalytic site on intracellular
CC distribution of the PDE for the preparation of a medicament. The present
CC sequence represents the amino acid sequence of a HSPDE4A4-E222G fusion
CC protein
XX
XX Sequence 893 AA;
XX
XX Query Match 99.8%; Score 1272; DB 4; Length 893;
XX Best Local Similarity 99.6%; Pred. No. 1.1e-121;
XX Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 MYSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
XX 655 MYSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 714
XX
XX 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
XX 715 LVTTLISYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 774
XX
XX 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
XX 775 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 834
XX
XX 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239
XX 835 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVIGFVTAAGITLGMDELYK 893
XX
XX RESULT 4
XX AAG65782
XX ID AAG65782 standard; protein; 1132 AA.
XX
XX AC AAG65782;
XX
XX XX
XX DT 07-JAN-2002 (first entry)
XX
XX DE Amino acid sequence of HSPDE4A4-E222G fusion protein.
XX
XX PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;
KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;
KW fusion protein.
XX
XX OS Homo sapiens.
XX Aequorea victoria.
XX
XX PN WO200179526-A2.
XX
```

```
PD 25-OCT-2001.
XX
XX 11-APR-2001; 2001WO-DK000264.
XX
XX 17-APR-2000; 2000DK-00000651.
XX 29-MAY-2000; 2000DK-00000849.
XX
XX (BIOI-) BIOIMAGE AS.
XX
XX Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;
XX Praestegaard M;
XX
XX WPI: 2001-611727/70.
XX N-PSDB; AA166853.
XX
XX Determining if a compound is a dislocator of PDE4 for identifying
XX compounds for treating CNS and inflammatory disease comprises identifying
XX compounds which remove PDE4 spots.
XX
XX Example 1; Page 162-167; 160pp; English.
XX
XX The invention relates to determining, if a compound, is a dislocator of
XX PDE4. The method comprises testing if the compound removes PDE4 spots,
XX which may optionally be induced by a Rolipram-like reference compound,
XX and testing if it inhibits the catalytic activity of the PDE4, where the
XX compound is a dislocator of PDE4, if it removes PDE spots and if it does
XX not inhibit the catalytic activity of PDE4. The method is useful for
XX identifying compounds useful for the treatment of diseases of the central
XX nervous system such as depression and for the treatment of inflammatory
XX disease such as joint inflammation, Crohn's disease, inflammatory bowel
XX disease, respiratory diseases, chronic obstructive pulmonary disease
XX (COPD), including asthma, chronic bronchitis, pulmonary emphysema,
XX endotoxemic shock, toxic shock syndrome, systemic lupus erythematosus,
XX psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV
XX infection. The use of a reagent that can mimic or reverse the effect of
XX the compound with affinity for the catalytic site on intracellular
XX distribution of the PDE for the preparation of a medicament. The present
XX sequence represents the amino acid sequence of a HSPDE4A4-E222G fusion
XX protein
XX
XX Sequence 1132 AA;
XX
XX Query Match 99.8%; Score 1272; DB 4; Length 1132;
XX Best Local Similarity 99.6%; Pred. No. 1.5e-121;
XX Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 MYSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60
XX 894 MYSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 953
XX
XX 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
XX 954 LVTTLISYGVQCFSRYPDHMKQHDFFKSAAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 1013
XX
XX 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
XX 1014 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 1073
XX
XX 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239
XX 1074 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMLVIGFVTAAGITLGMDELYK 1132
XX
XX RESULT 5
XX AAE17517
XX ID AAE17517 standard; protein; 239 AA.
XX
XX AC AAE17517;
XX
XX XX
XX DT 22-APR-2002 (first entry)
XX
XX DE Enhanced F64L jellyfish green fluorescent protein mutant.
XX
```

KW	Jellyfish; green fluorescent protein; GFP; protein redistribution; cellular function; genetic reporter; mutant; Stoke's shift; muten.	
XX		
OS	Aequorea victoria.	
OS	Synthetic.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 65	/note= "Wild type Phe substituted with Leu; This corresponds to position 64 in the wild type protein"
FT		
FT		
PN	WO200198338-A2.	
XX		
XX	27-DEC-2001.	
XX		
XX	18-JUN-2001; 2001WO-EP006848.	
XX		
PR	19-JUN-2000; 2000DK-00000953.	
PR	20-JUN-2000; 2000US-0212681P.	
PR	10-MAY-2001; 2001DK-00000739.	
PR	10-MAY-2001; 2001US-0290170P.	
XX		
PA	(BIOI-) BIOIMAGE AS.	
XX		
PI	Bjorn SP, Pagliaro L, Thastrup O;	
XX		
DR	WPI; 2002-098224/13.	
DR	N-PSDB; AAD28162.	
XX		
XX	Novel fluorescent protein in in vitro assay for measuring protein kinase activity or dephosphorylation activity, or for measuring protein redistribution, has a green fluorescent protein with F64L and E222G mutation.	
PT		
PT		
PT		
XX		
PS	Example 1; Page 35; 41pp; English.	
XX		
CC	The invention relates to a fluorescent protein derived from green fluorescent protein (GFP) or its analogue. The GFP containing mutations at F64L and E222G has a bigger compared to other GFP's making it very suitable for high throughput screening due to better resolution. The fluorescent protein is useful in in vitro assays for measuring protein kinase activity or dephosphorylation activity, or for measuring protein redistribution. The fluorescent protein is useful in studying cellular functions in living cells; as protein tags in transgenic animals, living and fixed cells; organelle tags, secretion marker and genetic reporter. The fluorescent protein is also useful as a cell or organelle integrity marker, a marker for changes in cell morphology, as transfection marker, and as a marker to be used in combination with fluorescence activated cell sorting (FACS). The novel proteins can also be used as reporters to monitor live or dead biomass of organisms, such as fungi. The fluorescent protein is also useful as markers in transcriptional and translational fusions for performing transposon vector mutagenesis and as a reporter for bacterial detection. Transposons encoding the fluorescent protein are useful for screening promoters and for tagging plasmids and chromosomes. The fluorescent protein engineered into the genome of a phage is useful for designing diagnostic tool. The present sequence is enhanced F64L jellyfish green fluorescent protein (GFP) mutant	
XX		
SQ	Sequence 239 AA;	
	Query Match	
	Best Local Similarity 99.2%; Score 1264; DB 5; Length 239;	
	Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;	
QY	1 MVSKEELFTGVPIVLELGDGVNGHKFVSVEGEGDATYKGLIKFKICTTCKLPVPMPT 60	
DB	1 MVSKEELFTGVPIVLELGDGVNGHKFVSVEGEGDATYKGLIKFKICTTCKLPVPMPT 60	
QY	61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAHPGYSVQERTIFFKDDGNYKTRAEVKEGDTL 120	
DB	61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAHPGYSVQERTIFFKDDGNYKTRAEVKEGDTL 120	
QY	121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIWADKQKNGIKVNFKIRHNIEDGSVQLA 180	

Db	121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIWADKQKNGIKVNFKIRHNIEDGSVQLA 180	
QY	181 DHYQNTPTGDPVLLPDPNHYLSTQSALSKDPNEKRDHMLVIGFTAAAGITLGMDELYK 239	
DB	181 DHYQNTPTGDPVLLPDPNHYLSTQSALSKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239	
	RESULT 6	
ID	AAB22882	
XX	AAB22882 standard; protein; 239 AA.	
AC	AAB22882;	
XX		
DT	10-JAN-2001 (first entry)	
XX		
DE	Enhanced green fluorescent protein (EGFP), SEQ ID NO:46.	
XX		
KW	Bioreporter protein; fusion protein; recognition site; cellular targeting sequence; cellular localisation; fluorescent protein; protease activity detection; toxin detection; cellular stress detection; drug discovery; cell based screening.	
XX		
OS	Aequorea victoria.	
OS	Synthetic.	
XX		
PN	WO2000050872-A2.	
XX		
PD	31-AUG-2000.	
XX		
XX	25-FEB-2000; 2000WO-US004794.	
XX		
PR	26-FEB-1999; 99US-0122152P.	
PR	08-MAR-1999; 99US-0123399P.	
PR	12-JUL-1999; 99US-00352171.	
XX		
XX	(CELL-) CELLOMICS INC.	
PA		
XX	Giuliano KA, Kapur R;	
PI		
XX	WPI; 2000-594086/56.	
DR	N-PSDB; AAA93373.	
XX		
XX	Automated cell-based characterization of toxin by contacting cells containing luminescent reporter molecules with test substance and analyzing optically.	
PT		
PT		
XX		
PS	Example 11; Fig 29A; 336pp; English.	
XX		
CC	The invention relates to systems, methods and reagents for cell-based screening or detection of compounds which affect particular biological functions. The methods of the invention utilise fluorescent bioreporter molecules which, when acted on by a compound of interest, cause an alteration in the cellular distribution of at least the fluorescent moiety. In one embodiment, the bioreporter comprises heat shock proteins (HSPs) fused to a fluorescent protein (e.g., jellyfish green fluorescent protein (GFP), or derivatives thereof). Such bioreporters are located in the cytoplasm, but on stress activation translocate to the nucleus. In another embodiment bioreporter proteins can be used to detect protease activity. Such protease bioreporter fusion proteins comprise one or more fluorescent proteins; a recognition signal which is cleaved by the protease; and at least one cellular localisation signal. The latter two components may be components of a single protein which is acted upon by the protease, or may be from heterologous sources. Due to the localisation signal, the bioreporter protein is localised to a particular region of the cell. Once acted on by the protease of interest, the fluorescent protein is cleaved from the localisation sequence, and is free to migrate to other locations within the cell. The presence of a second localisation signal attached to the fluorescent protein enables the fluorescent protein to be directed to a different cellular compartment after cleavage of the protease recognition sequence. The change in distribution of the fluorescent protein can be detected using imaging methods with a high degree of spatial resolution. The methods and	

CC biosensors of the invention can be used to investigate a wide range of  
 CC cellular activities and to screen compounds which modulate these  
 CC activities. Biosensors containing a recognition site for caspase, for  
 CC example, may be used for the screening of compounds which modulate  
 CC apoptosis, while biosensors containing other protease recognition sites  
 CC may be used for the detection of proteolytic toxins (such as anthrax  
 CC lethal factor). The method provides improved target validation and  
 CC candidate compound optimisation by combining many cell screening formats  
 CC with fluorescence-based molecular reagents and computer-based feature  
 CC extraction, data analysis and automation, resulting in increased quantity  
 CC and speed of data collection and faster evaluation of drug candidates.  
 CC Sequences AAB22881-B22885 represent fluorescent proteins which may be used  
 CC as components of biosensor fusion proteins of the invention  
 XX  
 SQ Sequence 239 AA;

Query Match 99.0%; Score 1261; DB 3; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 2.1e-121;  
 Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MVSKEBELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYKLTAKFICTTGKLPVPWPT 60  
 DB 1 MVSKEBELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYKLTAKFICTTGKLPVPWPT 60  
 QY 61 LVTLLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNVMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 QY 181 DHYQONTPIGDGVPLLPDNNHLSQTSALSKDPNEKRDHVMVLFVTAAGITLGMDELYK 239  
 DB 181 DHYQONTPIGDGVPLLPDNNHLSQTSALSKDPNEKRDHVMVLFVTAAGITLGMDELYK 239

RESULT 7  
 AAY54349  
 ID AAY54349 standard; protein; 239 AA.  
 XX  
 AC AAY54349;  
 XX  
 DT 06-APR-2000 (first entry)  
 DE Amino acid sequence of the mutant green fluorescent protein EGFP.  
 XX  
 XX Fluorescent protein; green fluorescent protein; emission intensity;  
 KW fluorescence; pH detection; pH sensor; EGFP.  
 OS Synthetic.  
 OS Aequorea victoria.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 65 /note= "wild type Phe substituted with Leu"  
 FT Misc-difference 66 /note= "wild type Ser substituted with Thr"  
 FT Misc-difference 232 /note= "wild type His substituted with Leu"  
 FT  
 XX  
 PN W0964592-A2.  
 XX  
 PD 16-DEC-1999.  
 XX  
 PF 08-JUN-1999; 99WO-US012850.  
 XX  
 PR 09-JUN-1998; 98US-00094359.  
 PR 13-OCT-1998; 98US-00172063.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 PA (UYOR-) UNIV OREGON STATE.  
 XX

PI Tsien RY, Llopis J, Wachter RM;  
 XX WPI: 2000-116540/10.  
 DR N-PSDB; AA245642.  
 XX New functional engineered green fluorescent proteins, used for measuring  
 PT the pH in biological samples and cells.  
 XX  
 PS Disclosure; Page 9; 89pp; English.  
 XX  
 CC The present sequence represents a functional engineered fluorescent  
 CC protein based on the Aequorea green fluorescent protein (GFP). The  
 CC emission intensity changes as pH varies between 5 and 10 of the present  
 CC protein are novel. The functional engineered fluorescent proteins show  
 CC reversible changes in fluorescence over physiological pH ranges. They can  
 CC be used for determining the pH of samples and cells. The polynucleotides  
 CC can also be used to produce transgenic animals. The fluorescent protein  
 CC pH sensors can be delivered to cells in the form of polynucleotides  
 CC encoding the protein sensor fused to a targeting signal. The targeting  
 CC signal directs the expression of the protein sensors to restricted cell  
 CC locations. This makes it possible to measure the pH of a precisely  
 CC defined cellular region or organelle  
 XX  
 SQ Sequence 239 AA;

Query Match 99.0%; Score 1261; DB 3; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 2.1e-121;  
 Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MVSKEBELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYKLTAKFICTTGKLPVPWPT 60  
 DB 1 MVSKEBELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYKLTAKFICTTGKLPVPWPT 60  
 QY 61 LVTLLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTLLTYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNVMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 QY 181 DHYQONTPIGDGVPLLPDNNHLSQTSALSKDPNEKRDHVMVLFVTAAGITLGMDELYK 239  
 DB 181 DHYQONTPIGDGVPLLPDNNHLSQTSALSKDPNEKRDHVMVLFVTAAGITLGMDELYK 239

RESULT 8  
 AAY79584  
 ID AAY79584 standard; peptide; 239 AA.  
 XX  
 AC AAY79584;  
 XX  
 DT 29-AUG-2000 (first entry)  
 DE EGFP signal domain.  
 XX  
 KW Protease; biosensor; EGFP; signal peptide; cell screening; assay;  
 KW analysis; drug discovery.  
 XX  
 OS Unidentified.  
 XX  
 PN W0200026408-A2.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 29-OCT-1999; 99WO-US025431.  
 XX  
 PR 30-OCT-1998; 98US-0106308P.  
 PR 26-MAY-1999; 99US-0136078P.  
 XX  
 PA (CELL-) CELLOMICS INC.  
 XX  
 PI Guiliano KA, Bright G, Olson K, Burroughs-Tencza S;

XX WPI; 2000-365644/31.  
DR N-PSDB; AA27573.  
XX Recombinant nucleic acid encoding a protease biosensor useful for  
PT fluorescence based cell and molecular biochemical assays for drug  
PT discovery comprising three operably linked nucleic acid sequences.  
XX  
XX Claim 14; Fig 29A; 218pp; English.  
XX  
CC The present sequence is that of the EGFP signal domain, which can be  
CC included in novel recombinant protease biosensors (PBs) of the invention.  
CC The PBs (see AA79638-54) comprise: a first domain (see AA79579-87)  
CC comprising at least 1 detectable polypeptide signal such as the present  
CC sequence; a second domain (see AA79588-622) comprising at least 1  
CC protease recognition site; and a third domain (see AA79623-37)  
CC comprising at least 1 reactant target sequence. A recombinant nucleic  
CC acid (see AA27627-43) encoding the PB, an expression vector, and a  
CC genetically engineered host cell are also claimed. A claimed method for  
CC identifying compounds that modify protease activity in a cell involves  
CC contacting a host cell that possesses the recombinant PB with a test  
CC compound, and determining the PB distribution in the host cell, where  
CC changes in the distribution of the PB are correlated with modification of  
CC protease activity by the test compound. Claimed kits for identifying  
CC compounds that modify protease activity in a host cell include the  
CC recombinant nucleic acid, or the recombinant PB, or the vector, or the  
CC host cell. The PB is useful in high content screens to detect in vivo  
CC activation of enzymatic activity, and to identify specific activity based  
CC on cleavage of a known recognition motif  
XX  
XX Sequence 239 AA;  
SQ

Query Match 99.0%; Score 1261; DB 3; Length 239;  
Best Local Similarity 98.7%; Pred. No. 2.1e-121;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYCKLTGKLPVWPWT 60  
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYCKLTGKLPVWPWT 60  
QY 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRLEKGIIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
DB 121 VNRLEKGIIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLIGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLIGFVTAAGITLGMDELYK 239

RESULT 9  
AAB50804  
ID AAB50804 standard; protein; 239 AA.  
XX  
XX AAB50804;  
AC  
DT 14-MAR-2001 (first entry)  
XX  
XX Jellyfish GFP mutant EGFP.  
DE  
KW Aequorea victoria; jellyfish; fluorescent protein indicator;  
XX green fluorescent protein; GFP; linker moiety; sensor;  
KW calmodulin-binding domain; mutant; mutin.  
XX  
XX Aequorea victoria.  
OS  
XX WO200071565-A2.  
XX  
XX 30-NOV-2000.  
XX

PF 17-MAY-2000; 2000WO-US013684.  
XX  
XX 21-MAY-1999; 99US-00316919.  
PR 21-MAY-1999; 99US-00316920.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
XX  
XX Tsien RV, Baird GA;  
PI  
XX WPI; 2001-032017/04.  
DR N-PSDB; AAC90488.  
XX  
XX Novel fluorescent proteins comprising a sensor protein inserted into  
PT them, useful for measuring the response of a sensor biological, chemical,  
PT electrical or physiological parameter in vivo or in vitro.  
XX  
XX Disclosure; Page 24; 94pp; English.  
PS  
XX The present sequence is a fluorescent protein used in the construction of  
CC a fluorescent protein indicator. The indicator comprises a sensor  
CC polypeptide that is responsive to a chemical, biological, electrical or  
CC physiological parameter, and a fluorescence protein functional group. The  
CC sensor polypeptide is operatively inserted into the fluorescent moiety.  
CC The fluorescent indicator is useful for detecting the presence of a  
CC response inducing member in a sample. The method involves contacting the  
CC sample with the indicator and detecting a change in fluorescence, in  
CC which a change is indicative of the effect of the parameter on the sensor  
CC polypeptide. The novel fluorescent proteins are advantageous due to their  
CC reduced size as compared to the FRET (fluorescence resonance energy  
CC transfer)-based sensors  
XX  
XX Sequence 239 AA;  
SQ

Query Match 99.0%; Score 1261; DB 4; Length 239;  
Best Local Similarity 98.7%; Pred. No. 2.1e-121;  
Matches 236; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYCKLTGKLPVWPWT 60  
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYCKLTGKLPVWPWT 60  
QY 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRLEKGIIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
DB 121 VNRLEKGIIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLIGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLIGFVTAAGITLGMDELYK 239

RESULT 10  
AAB85900  
ID AAB85900 standard; protein; 239 AA.  
XX  
XX AAB85900;  
AC  
DT 30-NOV-2001 (first entry)  
XX  
XX A. victoria green fluorescent protein (GFP) and linker sequence.  
DE  
XX Melanin concentrating hormone receptor; MCHR; MCH; chimeric; fusion;  
KW green fluorescent polypeptide; orexinergic; anabolic; food intake; GFP;  
KW green fluorescent protein.  
XX  
XX Synthetic.  
OS Aequorea victoria.  
XX  
XX WO200168706-A1.  
XX



```
PF 04-AUG-2000; 2000JP-00237165.
XX
PR 04-AUG-2000; 2000JP-00237165.
XX
XX (RIKA ) RIKAGAKU KENKYUSHO.
PA
XX WPI; 2002-299190/34.
XX N-PSDB; ABL40628.
XX
XX A gene encoding cyan-green fluorescent protein.
XX
XX Example; Page 14; 20pp; Japanese.
XX
XX The invention relates to a gene encoding proteins having cyan-green
XX fluorescence characteristic and having a function of showing stable
XX fluorescence characteristic in acid region. A method for the preparation
XX of a cyan-green fluorescent protein is provided which involves a
XX transformant transformed by a recombinant vector comprising the gene,
XX where the transformant is cultured and the protein is collected from the
XX culture. The present sequence represents the A. victoria green
XX fluorescent protein (EGFP)
XX
XX Sequence 239 AA;
XX
XX Query Match 99.0%; Score 1261; DB 5; Length 239;
XX Best Local Similarity 98.7%; Pred. No. 2.1e-121;
XX Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTLPKICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTLPKICTTGKLPVWPWT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLELKGIDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRLELKGIDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLTGVTAAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLTGVTAAAGITLGMDELYK 239
RESULT 13
ABG94444
ID ABG94444 standard; protein; 239 AA.
XX
AC ABG94444;
XX
XX 27-NOV-2002 (first entry)
XX
XX Protease biosensor signal sequence #6.
XX
XX Detection; classification; identification; toxin detection; protease;
XX ADP-ribosylating toxin; cytotoxic phospholipase; exfoliative toxin;
XX toxic threat agent.
XX
XX Synthetic.
XX
XX US6413959-B1.
XX
XX 09-JUL-2002.
XX
XX 25-FEB-2000; 2000US-00513783.
XX
XX 27-FEB-1997; 97US-00810983.
XX
XX 27-FEB-1998; 98US-00031271.
XX
XX 26-FEB-1999; 99US-0122152P.
XX
XX 08-MAR-1999; 99US-0123399P.
XX
XX 12-JUL-1999; 99US-00352171.
XX
XX 31-AUG-1999; 99US-0151797P.
XX
PR 17-SEP-1999; 99US-00398965.
PR 29-OCT-1999; 99US-00430656.
PR 01-DEC-1999; 99US-0168408P.
XX
XX (GIUL/) GIULIANO K.
XX (KAPU/) KAPUR R.
XX
XX Giuliano K, Kapur R;
XX
XX WPI; 2002-634730/68.
XX N-PSDB; ABS711491.
XX
XX Automated cell-based toxin detection, classification, and/or
XX identification by treating cells involves use of three classes of
XX luminescent reporter molecules such as detectors, classifiers or
XX identifiers.
XX
XX Example 10; Fig 29A; 214pp; English.
XX
XX The invention describes methods of automated detection, classification
XX and identification comprising treating cells containing luminescent
XX reporter molecules (I) in array of locations with a test substance, where
XX (I) are detectors, classifiers or identifiers, imaging cells in each
XX location to obtain luminescent signals and converting optical information
XX into digital data to interpret presence of toxins in the test substance.
XX The method are useful for detection of toxins chosen from proteases, ADP-
XX ribosylating toxins, cytotoxic phospholipases, and exfoliative toxins.
XX Three classes of cell-based luminescent reporter molecules such as
XX detectors, classifiers and identifiers are described and serve as
XX reporters of toxic threat agents. The first two levels of
XX characterisation ensure a rapid readout of toxin class without
XX sacrificing the ability to detect many new mutant toxins or dissect
XX several complex mixtures of known toxins. This is the amino acid sequence
XX of a protease biosensor related signal sequence used in the cell-based
XX screening system
XX
XX Sequence 239 AA;
XX
XX Query Match 99.0%; Score 1261; DB 5; Length 239;
XX Best Local Similarity 98.7%; Pred. No. 2.1e-121;
XX Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTLPKICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTLPKICTTGKLPVWPWT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRLELKGIDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRLELKGIDFKEDGNILGHKLEYNHNHVMADKQKNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLTGVTAAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLTGVTAAAGITLGMDELYK 239
RESULT 14
AAE14599
ID AAE14599 standard; protein; 239 AA.
XX
XX AAE14599;
XX
XX 31-MAY-2002 (first entry)
XX
XX Aequorea victoria enhanced green fluorescent protein.
XX
XX Mutagenesis; enhanced green fluorescent protein; EGFP; mutant; mutein.
XX
XX Aequorea victoria.
XX
XX Synthetic.
```



XX Key Location/Qualifiers  
 FH Misc-difference 1. .3  
 FT /note= "Wild-type GFP Met-Ser are replaced with Met-Val-Ser"  
 FT  
 FT Misc-difference 65  
 FT /note= "GFP phe64 is replaced by Leu"  
 FT Misc-difference 66  
 FT /note= "GFP Ser65 is replaced by Thr"  
 XX  
 XX EP1178109-A1.  
 XX  
 XX 06-FEB-2002.  
 XX  
 XX 03-AUG-2001; 2001BP-00306650.  
 XX  
 XX 04-AUG-2000; 2000JP-00237166.  
 PR  
 XX (RIKE ) RIKEN KK.  
 PA  
 XX Miyawaki A, Sawano A;  
 PI WPI; 2002-208112/27.  
 DR N-PSDB; AAD27910.  
 DR  
 XX  
 XX Method for mutagenesis, e.g. for introducing certain or random mutations at certain sites of the nucleotide sequence, comprises synthesizing a mutated strand and a complementary strand by use of megaprimer.  
 PT  
 PT  
 PT  
 XX  
 PS Example 1; Page 13-14; 31pp; English.  
 XX  
 CC The invention relates to a method for mutagenesis that comprises synthesising a mutated strand and a complementary strand by use of megaprimer. The method basically comprises a DNA synthesis in which one or more primers that have a nucleotide sequence containing at least one mutation and a phosphorylated 5'-terminus are annealed to a template DNA and then subjected to an elongation reaction using a thermostable high-fidelity DNA polymerase, after which the phosphorylated 5'-terminus and the elongated terminus are ligated by means of a thermostable DNA ligase to synthesise a circular DNA containing the primers; a digestion in which the step of DNA synthesis is repeated several times to amplify the DNA containing the primers and then, at least DNAs other than the amplified circular DNA are digested into several fragments; and a double-stranded DNA synthesis in which, with the several fragments obtained in the step of digestion as megaprimers, the megaprimers are annealed to the circular DNA synthesised above, followed by an elongation reaction performed using the thermostable high-fidelity DNA polymerase. The method is useful for mutagenesis, particularly for introducing certain mutations at certain sites of a nucleotide sequence, or for introducing random mutations at certain sites of the nucleotide sequence. The present method is simple, speedy, economical and widely applicable. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used for mutagenesis in an exemplification of the invention. The EGFP is derived by altering the green fluorescent protein (GFP) sequence of Aequorea victoria  
 XX  
 XX Sequence 239 AA;  
 SQ  
 Query Match 99.0%; Score 1261; DB 5; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 2.1e-121;  
 Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MYSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTCKLPVPWPT 60  
 DB 1 MYSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTCKLPVPWPT 60  
 QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTTLTYGVCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGDIFKEDGNILGHKLEYNYNHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
 DB 121 VNRIELKGDIFKEDGNILGHKLEYNYNHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
 QY 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239  
 DB 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239

QY 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239  
 DB 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239  
 RESULT 15  
 AAE34958  
 ID AAE34958 standard; protein; 239 AA.  
 XX  
 AC AAE34958;  
 XX  
 DT 28-MAY-2003 (first entry)  
 XX  
 DE Aequorea victoria enhanced green fluorescent protein (EGFP).  
 XX  
 XX Phosphorylation indicator; fluorescent protein; detection; phosphatase;  
 KW kinase; enhanced green fluorescent protein; EGFP.  
 XX  
 OS Aequorea victoria.  
 XX  
 PN WO200295058-A2.  
 XX  
 PD 28-NOV-2002.  
 XX  
 PF 24-MAY-2002; 2002WO-US016955.  
 XX  
 PR 24-MAY-2001; 2001US-00865291.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 XX Tsien RY, Ting AY, Zhang J;  
 XX WPI; 2003-148474/14.  
 DR N-PSDB; AAD53428.  
 XX  
 XX Novel chimeric phosphorylation indicators, useful for detecting kinase/phosphatase in samples, has donor molecule, phosphorylatable domain, phosphoaminoacid binding domain, and acceptor molecule, in operative linkage.  
 PT  
 XX Disclosure; Col 56-57; 38pp; English.  
 PS  
 CC The present invention relates to chimeric phosphorylation indicators comprising a phosphorylation polypeptide and a fluorescent protein or in operative linkage, a donor molecule, a phosphorylatable domain, a phosphoaminoacid binding domain (PAABD) and an acceptor molecule. The phosphorylation indicators of the invention are useful for detecting kinases or phosphatases in a biological sample. They are also useful in high throughput analysis e.g. for detecting a kinase inhibitor or phosphatase inhibitor. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used in the invention  
 XX  
 SQ Sequence 239 AA;  
 Query Match 99.0%; Score 1261; DB 6; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 2.1e-121;  
 Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MYSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTCKLPVPWPT 60  
 DB 1 MYSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTCKLPVPWPT 60  
 QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTTLTYGVCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGDIFKEDGNILGHKLEYNYNHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
 DB 121 VNRIELKGDIFKEDGNILGHKLEYNYNHNVIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
 QY 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239  
 DB 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239

us-09-887-784-222i.rag

Sun Jun 27 18:27:31 2004

Db 181 DHYQONTPIGDGFVLLPDNHYLSTOSALS KDPNEKRDHNVLLFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 15:54:11  
Job time : 47.1111 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:55 ; Search time 12.7778 Seconds  
(without alignments)  
965.630 Million cell updates/sec

Title: US-09-887-784-222i  
Perfect score: 1274  
Sequence: 1 MWSKGBELFTGVVPIVLVELD.....VLIGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A COMB.pep:\*  
2: /cgn2\_6/prodata/2/iaa/5B COMB.pep:\*  
3: /cgn2\_6/prodata/2/iaa/6A COMB.pep:\*  
4: /cgn2\_6/prodata/2/iaa/6B COMB.pep:\*  
5: /cgn2\_6/prodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/iaa/backfileesl.pep:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1261	99.0	239	3	US-09-172-063-3
2	1261	99.0	239	4	US-09-513-783A-46
3	1261	99.0	239	4	US-09-316-919-4
4	1261	99.0	239	4	US-09-602-641-3
5	1261	99.0	239	4	US-09-920-922-2
6	1261	99.0	281	3	US-09-062-102-1
7	1261	99.0	281	4	US-09-364-946-1
8	1261	99.0	294	4	US-09-513-783A-2
9	1261	99.0	323	3	US-09-172-063-21
10	1261	99.0	323	4	US-09-602-641-21
11	1261	99.0	364	3	US-09-085-305-6
12	1261	99.0	379	4	US-09-417-197-129
13	1261	99.0	434	4	US-09-800-170-48
14	1261	99.0	442	4	US-09-417-197-127
15	1261	99.0	459	4	US-09-513-783A-170
16	1261	99.0	544	4	US-09-417-197-113
17	1261	99.0	544	4	US-09-417-197-115
18	1261	99.0	604	4	US-09-417-197-59
19	1261	99.0	605	4	US-09-417-197-41
20	1261	99.0	606	4	US-09-417-197-65
21	1261	99.0	607	4	US-09-417-197-47
22	1261	99.0	630	4	US-09-417-197-63
23	1261	99.0	631	4	US-09-417-197-39
24	1261	99.0	633	4	US-09-417-197-45
25	1261	99.0	635	4	US-09-417-197-125
26	1261	99.0	642	2	US-08-818-253-2
27	1261	99.0	642	2	US-08-818-253-6

ALIGNMENTS

RESULT 1  
US-09-172-063-3  
; Sequence 3, Application US/09172063  
; Patent No. 6150176  
; GENERAL INFORMATION:  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Miyawaki, Atsushi  
; APPLICANT: Liopis, Juan  
; APPLICANT: Wachter, Rebekka M.  
; APPLICANT: Remington, S. James  
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
; FILE REFERENCE: 07257/071001  
; CURRENT APPLICATION NUMBER: US/09/172,063  
; EARLIER FILING DATE: 1998-10-13  
; EARLIER APPLICATION NUMBER: 09/094,359  
; EARLIER FILING DATE: 1998-06-09  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: EGFP  
US-09-172-063-3

Query Match 99.0%; Score 1261; DB 3; Length 239;  
Best Local Similarity 98.7%; Pred. No. 1.9e-127;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MWSKGBELFTGVVPIVLVELDGVNGHKFSVSGEGDATYVKLTAKFKICTTGKLPVPWPT 60  
Db 1 MWSKGBELFTGVVPIVLVELDGVNGHKFSVSGEGDATYVKLTAKFKICTTGKLPVPWPT 60  
QY 61 LVTTLSYGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNVKTAEVKFEGDTL 120  
Db 61 LVTTLSYGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNVKTAEVKFEGDTL 120  
QY 121 VRIELKIDFKEDGNILGHKLEYNYSNNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
Db 121 VRIELKIDFKEDGNILGHKLEYNYSNNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPTGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239  
Db 181 DHYQONTPTGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239

Sequence 2, Appli  
Sequence 6, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 75, Appli  
Sequence 51, Appli  
Sequence 139, App  
Sequence 176, App  
Sequence 141, App  
Sequence 143, App  
Sequence 77, Appli  
Sequence 178, App  
Sequence 53, Appli  
Sequence 61, Appli  
Sequence 43, Appli  
Sequence 117, App  
Sequence 119, App

Sun Jun 27 18:27:31 2004

```
RESULT 2
US-09-513-783A-46
; Sequence 46, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 46
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGFP
US-09-513-783A-46
Query Match          99.0%; Score 1261; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.9e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTILKFICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTILKFICTTGKLPVWPWT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDYL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDYL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLIGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLIGFVTAAGITLGMDELYK 239
RESULT 3
US-09-316-919-4
; Sequence 49, Application US/09316919
; Patent No. 6469154
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: 07257/073001
; CURRENT APPLICATION NUMBER: US/09/316,919
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-316-919-4
Query Match          99.0%; Score 1261; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.9e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTILKFICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTILKFICTTGKLPVWPWT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDYL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDYL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLIGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLIGFVTAAGITLGMDELYK 239
RESULT 4
US-09-602-641-3
; Sequence 3, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-602-641-3
Query Match          99.0%; Score 1261; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.9e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTILKFICTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDGYGKLTILKFICTTGKLPVWPWT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDYL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDYL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLIGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLIGFVTAAGITLGMDELYK 239
RESULT 5
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. 6673610
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; 
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; US-09-920-922-2

Query Match          99.0%; Score 1261; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.9e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSGBELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MVSGBELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 LVTTLYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 6
US-09-062-102-1
; Sequence 1, Application US/09062102
; Patent No. 6130313
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangliang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100
; CURRENT APPLICATION NUMBER: US/09/062,102
; CURRENT FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: US 60/060,855
; EARLIER FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6130313
US-09-062-102-1

Query Match          99.0%; Score 1261; DB 3; Length 281;
Best Local Similarity 98.7%; Pred. No. 2.4e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSGBELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MVSGBELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 LVTTLYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
RESULT 7
US-09-364-946-1
; Sequence 1, Application US/09364946
; Patent No. 6306600
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangliang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100CJP/D2
; CURRENT APPLICATION NUMBER: US/09/364,946
; CURRENT FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: US 09/191,233
; EARLIER FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6306600
US-09-364-946-1

Query Match          99.0%; Score 1261; DB 4; Length 281;
Best Local Similarity 98.7%; Pred. No. 2.4e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSGBELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MVSGBELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 LVTTLYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 8
US-09-513-783A-2
; Sequence 2, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GFP-DEVD-Annexin II construct
US-09-513-783A-2

Query Match          99.0%; Score 1261; DB 4; Length 294;
Best Local Similarity 98.7%; Pred. No. 2.6e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSGBELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Db 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDGYGKLTGKLFCTTGKLPVWPWT 60  
QY 61 LVTTLTSGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120  
Db 61 LVTTLTSGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120  
QY 121 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
Db 121 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
QY 181 DHYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMLVIGFVTAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMLVIGFVTAAGITLGMDELYK 239

RESULT 9  
US-09-172-663-21  
; Sequence 21, Application US/09172063  
; Patent No. 6150176  
; GENERAL INFORMATION:  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Miyawaki, Atsushi  
; APPLICANT: Llopis, Juan  
; APPLICANT: Wachter, Rebekka M.  
; APPLICANT: Remington, S. James  
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
; FILE REFERENCE: 07257/071001  
; CURRENT FILING DATE: 1998-10-13  
; EARLIER APPLICATION NUMBER: 09/094,359  
; EARLIER FILING DATE: 1998-06-09  
; SOFTWARE: Fast-SEQ for Windows Version 4.0  
; SEQ ID NO 21  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: GT-EGFP  
US-09-172-663-21

Query Match 99.0%; Score 1261; DB 3; Length 323;  
Best Local Similarity 98.7%; Pred. No. 3e-127;  
Matches 236; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDGYGKLTGKLFCTTGKLPVWPWT 60  
Db 85 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDGYGKLTGKLFCTTGKLPVWPWT 144  
QY 61 LVTTLTSGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120  
Db 145 LVTTLTSGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 204  
QY 121 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
Db 205 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 264  
QY 181 DHYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMLVIGFVTAAGITLGMDELYK 239  
Db 265 DHYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMLVIGFVTAAGITLGMDELYK 323

RESULT 10  
US-09-602-641-21  
; Sequence 21, Application US/09602641  
; Patent No. 6608189  
; GENERAL INFORMATION:  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Miyawaki, Atsushi  
; APPLICANT: Llopis, Juan

; APPLICANT: Wachter, Rebekka M.  
; APPLICANT: Remington, S. James  
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
; FILE REFERENCE: 07257/071001  
; CURRENT APPLICATION NUMBER: US/09/602,641  
; CURRENT FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 09/172,063  
; PRIOR FILING DATE: 1998-10-13  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: Fast-SEQ for Windows Version 4.0  
; SEQ ID NO 21  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: GT-EGFP  
US-09-602-641-21

Query Match 99.0%; Score 1261; DB 4; Length 323;  
Best Local Similarity 98.7%; Pred. No. 3e-127; 1; Indels 0; Gaps 0;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDGYGKLTGKLFCTTGKLPVWPWT 60  
Db 85 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDGYGKLTGKLFCTTGKLPVWPWT 144  
QY 61 LVTTLTSGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 120  
Db 145 LVTTLTSGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNGYKTRAEVKFEGDTL 204  
QY 121 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
Db 205 VNRLELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 264  
QY 181 DHYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMLVIGFVTAAGITLGMDELYK 239  
Db 265 DHYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMLVIGFVTAAGITLGMDELYK 323

RESULT 11  
US-09-085-305-6  
; Sequence 6, Application US/09085305  
; Patent No. 6191269  
; GENERAL INFORMATION:  
; APPLICANT: Pollock, Allan  
; APPLICANT: Lovett, David H.  
; APPLICANT: Turck, Johanna  
; TITLE OF INVENTION: Selective Induction of Apoptosis in  
; TITLE OF INVENTION: Malignant Cancer Cells by Delivery of N-Terminal  
; TITLE OF INVENTION: Interleukin-1-Alpha Pro-Piece Polypeptide  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Bozicevic & Reed, LLP  
; STREET: 285 Hamilton Ave, Suite 200  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: Fast-SEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/085,305  
; FILING DATE: 29-MAY-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Francis, Carol L  
REGISTRATION NUMBER: 36,513  
REFERENCE/DOCKET NUMBER: 6510/102US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-327-3400  
TELEFAX: 650-327-3231  
TELEX:  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 364 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-085-305-6

Query Match 99.0%; Score 1261; DB 3; Length 364;  
Best Local Similarity 98.7%; Pred. No. 3.6e-127;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60  
Db 126 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 185

QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKFSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 186 LVTTLISYGVQCFSRYPDHMKQHDFFKFSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 245

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
Db 246 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGVSQVLA 305

QY 181 DHYQONTPIGDPVLLPDNHNLSQTSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239  
Db 306 DHYQONTPIGDPVLLPDNHNLSQTSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 364

RESULT 12  
US-09-417-197-129  
Sequence 129, Application US/09417197  
Patent No. 6518021  
GENERAL INFORMATION:  
APPLICANT: Ole THASTRUP, et al.  
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An  
FILE OF INVENTION: On A Cellular Response  
FILE REFERENCE: 3759-0110P  
CURRENT APPLICATION NUMBER: US/09/417,197  
CURRENT FILING DATE: 1999-10-07  
NUMBER OF SEQ ID NOS: 143  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 129  
LENGTH: 379  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: actin-binding-domain-EGFP fusion  
US-09-417-197-129

Query Match 99.0%; Score 1261; DB 4; Length 379;  
Best Local Similarity 98.7%; Pred. No. 3.8e-127;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60  
Db 141 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 200

QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKFSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 201 LVTTLISYGVQCFSRYPDHMKQHDFFKFSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 260

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
Db 261 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGVSQVLA 320

QY 181 DHYQONTPIGDPVLLPDNHNLSQTSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239  
Db 321 DHYQONTPIGDPVLLPDNHNLSQTSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 379

RESULT 13  
US-09-800-170-48  
Sequence 48, Application US/09800170  
Patent No. 6481667  
GENERAL INFORMATION:  
APPLICANT: Kinsella, Todd  
TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES  
FILE REFERENCE: A-68614-1/DJB/RMS/RMK  
CURRENT APPLICATION NUMBER: US/09/800,170  
CURRENT FILING DATE: 2001-08-28  
PRIOR APPLICATION NUMBER: US 60/187,130  
PRIOR FILING DATE: 2000-03-06  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 48  
LENGTH: 434  
TYPE: PRT  
ORGANISM: Synechocystis PCC6803  
US-09-800-170-48

Query Match 99.0%; Score 1261; DB 4; Length 434;  
Best Local Similarity 98.7%; Pred. No. 4.7e-127;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60  
Db 196 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 255

QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKFSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 256 LVTTLISYGVQCFSRYPDHMKQHDFFKFSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 315

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
Db 316 VNRIELKGIDFKEDGNILGHKLEYNHNVSIMADKQNGIKVNFKIRHNIEDGVSQVLA 375

QY 181 DHYQONTPIGDPVLLPDNHNLSQTSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239  
Db 376 DHYQONTPIGDPVLLPDNHNLSQTSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK 434

RESULT 14  
US-09-417-197-127  
Sequence 127, Application US/09417197  
Patent No. 6518021  
GENERAL INFORMATION:  
APPLICANT: Ole THASTRUP, et al.  
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An  
FILE OF INVENTION: On A Cellular Response  
FILE REFERENCE: 3759-0110P  
CURRENT APPLICATION NUMBER: US/09/417,197  
CURRENT FILING DATE: 1999-10-07  
NUMBER OF SEQ ID NOS: 143  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 127  
LENGTH: 442  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: EGFP-RhoA fusion  
US-09-417-197-127

Query Match 99.0%; Score 1261; DB 4; Length 442;  
Best Local Similarity 98.7%; Pred. No. 4.8e-127;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKHKFSVSGEGDATYGKLTAKFICTTGKLPVPWPT 60

Db 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDGYGKLTFLKFCITCTGKLPVPWPT 60  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRLEKGIIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180  
Db 121 VNRLEKGIIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239

RESULT 15  
US-09-513-783A-170  
; Sequence 170, Application US/09513783A  
; Patent No. 6416959  
; GENERAL INFORMATION:  
; APPLICANT: Giuliani, Kenneth A.  
; APPLICANT: Kapur, Ravi  
; TITLE OF INVENTION: A System for Cell Based Screening  
; FILE REFERENCE: 97-022-L1  
; CURRENT APPLICATION NUMBER: US/09/513,783A  
; CURRENT FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 170  
; LENGTH: 459  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP27  
US-09-513-783A-170

Query Match 99.0%; Score 1261; DB 4; Length 459;  
Best Local Similarity 98.7%; Pred. No. 5.1e-127; Indels 0; Gaps 0;  
Matches 236; Conservative 2; Mismatches 1;  
QY 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDGYGKLTFLKFCITCTGKLPVPWPT 60  
Db 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDGYGKLTFLKFCITCTGKLPVPWPT 60  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRLEKGIIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180  
Db 121 VNRLEKGIIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:04:03  
Job time : 12.7778 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:54:20 ; Search time 34.7778 Seconds  
(without alignments)  
1940.117 Million cell updates/sec

Title: US-09-887-784-2221

Perfect score: 1274

Sequence: 1 MYSKGEELFTGVVPIVLVELD.....VLIGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pdb.p\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pdb.p\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pdb.p\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pdb.p\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pdb.p\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pdb.p\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pdb.p\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pdb.p\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pdb.p\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pdb.p\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pdb.p\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pdb.p\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pdb.p\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pdb.p\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pdb.p\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pdb.p\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pdb.p\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pdb.p\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1272	99.8	239	9	US-09-887-784-4
2	1272	99.8	239	12	US-10-296-953-4
3	1272	99.8	239	14	US-10-270-223-6
4	1272	99.8	893	14	US-10-257-909A-30
5	1272	99.8	1132	14	US-10-257-909A-32
6	1264	99.2	239	9	US-09-887-784-2
7	1264	99.2	239	12	US-10-296-953-2
8	1261	99.0	239	9	US-09-920-922-2
9	1261	99.0	239	9	US-09-999-745-4
10	1261	99.0	239	10	US-09-866-538-4
11	1261	99.0	239	10	US-09-797-496B-2
12	1261	99.0	239	10	US-09-794-308-4
13	1261	99.0	239	10	US-09-865-291-4
14	1261	99.0	239	12	US-10-457-982-3
15	1261	99.0	239	14	US-10-121-258-13

16	1261	99.0	239	14	US-10-221-461-7	Sequence 7, Appli
17	1261	99.0	239	14	US-10-100-957A-46	Sequence 46, Appli
18	1261	99.0	239	14	US-10-177-390-2	Sequence 2, Appli
19	1261	99.0	239	14	US-10-338-411-3	Sequence 3, Appli
20	1261	99.0	239	15	US-10-370-570-4	Sequence 4, Appli
21	1261	99.0	239	15	US-10-389-640-3	Sequence 3, Appli
22	1261	99.0	259	14	US-10-314-861-11	Sequence 11, Appli
23	1261	99.0	281	12	US-09-931-232-1	Sequence 1, Appli
24	1261	99.0	288	14	US-10-314-861-37	Sequence 37, Appli
25	1261	99.0	293	14	US-10-314-861-35	Sequence 35, Appli
26	1261	99.0	294	14	US-10-100-957A-2	Sequence 2, Appli
27	1261	99.0	295	14	US-10-314-861-39	Sequence 39, Appli
28	1261	99.0	299	14	US-10-314-861-33	Sequence 33, Appli
29	1261	99.0	305	14	US-10-314-861-31	Sequence 31, Appli
30	1261	99.0	308	14	US-10-033-717-35	Sequence 35, Appli
31	1261	99.0	311	14	US-10-314-861-29	Sequence 29, Appli
32	1261	99.0	320	14	US-10-338-411-11	Sequence 11, Appli
33	1261	99.0	323	15	US-10-389-640-7	Sequence 7, Appli
34	1261	99.0	323	14	US-10-338-411-7	Sequence 13, Appli
35	1261	99.0	323	15	US-10-389-640-13	Sequence 13, Appli
36	1261	99.0	323	15	US-10-389-640-5	Sequence 5, Appli
37	1261	99.0	345	14	US-10-338-411-5	Sequence 9, Appli
38	1261	99.0	346	14	US-10-338-411-9	Sequence 9, Appli
39	1261	99.0	346	15	US-10-389-640-9	Sequence 9, Appli
40	1261	99.0	359	14	US-10-033-717-33	Sequence 33, Appli
41	1261	99.0	359	14	US-10-033-717-34	Sequence 34, Appli

ALIGNMENTS

RESULT 1

US-09-887-784-4  
; Sequence 4, Application US/09887784  
; Patent No. US20020177189A1  
; GENERAL INFORMATION:  
; APPLICANT: BUORN, Sara et al  
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS  
; FILE REFERENCE: 3759-0115P  
; CURRENT APPLICATION NUMBER: US/09/887,784  
; CURRENT FILING DATE: 2001-06-19  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 4  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequoria Victoria  
US-09-887-784-4

Query Match 99.8%; Score 1272; DB 9; Length 239;  
Best Local Similarity 99.6%; Pred. No. 2.2e-124;  
Matches 238; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY	1	MYSKGEELFTGVVPIVLVELDGVNKGKFSVSGEGDATYVKLTAKFICTTGLPVPWPT	60
DB	1	MYSKGEELFTGVVPIVLVELDGVNKGKFSVSGEGDATYVKLTAKFICTTGLPVPWPT	60
QY	61	LVTTLISYGVQCFRSRYPDHMKQHDFFKSAWPEGVVQERTIFFKDDGNVKTAEVKFEGDTL	120
DB	61	LVTTLISYGVQCFRSRYPDHMKQHDFFKSAWPEGVVQERTIFFKDDGNVKTAEVKFEGDTL	120
QY	121	VNRIELKIDFDEKDGNGILGHKLEYNYNHNVYIMADQKNGIKVNFKIRHNIEDGSVOLA	180
DB	121	VNRIELKIDFDEKDGNGILGHKLEYNYNHNVYIMADQKNGIKVNFKIRHNIEDGSVOLA	180
QY	181	DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK	239
DB	181	DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVIGFVTAAGITLGMDELYK	239



```
/ SEQ ID NO 32
/ LENGTH: 1132
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-32

Query Match
Best Local Similarity 99.8%; Score 1272; DB 14; Length 1132;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
DB 894 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 953
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 954 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 1013
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 1014 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 1073
QY 181 DHYQONTPIGDGPVLLPDNHHYLSQSALSQKDPNEKRDHVMVLCFVTAAGITTLGMDLEYK 239
DB 1074 DHYQONTPIGDGPVLLPDNHHYLSQSALSQKDPNEKRDHVMVLCFVTAAGITTLGMDLEYK 1132

RESULT 6
US-09-887-784-2
/ Sequence 2, Application US/09887784
/ Patent No. US20020177189A1
/ GENERAL INFORMATION:
/ APPLICANT: BJORN, Sara et al
/ TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
/ FILE REFERENCE: 3759-0115P
/ CURRENT APPLICATION NUMBER: US/09/887,784
/ CURRENT FILING DATE: 2001-06-19
/ NUMBER OF SEQ ID NOS: 24
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 2
/ LENGTH: 239
/ TYPE: PRT
/ ORGANISM: Aequorea Victoria
US-09-887-784-2

Query Match
Best Local Similarity 99.2%; Score 1264; DB 9; Length 239;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHHYLSQSALSQKDPNEKRDHVMVLCFVTAAGITTLGMDLEYK 239
DB 181 DHYQONTPIGDGPVLLPDNHHYLSQSALSQKDPNEKRDHVMVLCFVTAAGITTLGMDLEYK 239

RESULT 7
US-10-296-953-2
/ Sequence 2, Application US/10296953
/ Publication No. US20040072995A1
/ GENERAL INFORMATION:
```

```
/ APPLICANT: BJORN, SARA P.
/ APPLICANT: PAGLIARO, LEN
/ APPLICANT: THASTRUP, OLE
/ TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
/ FILE REFERENCE: PL0095
/ CURRENT APPLICATION NUMBER: US/10/296,953
/ CURRENT FILING DATE: 2002-11-26
/ PRIOR APPLICATION NUMBER: PA 2000 00953
/ PRIOR FILING DATE: 2000-06-19
/ PRIOR APPLICATION NUMBER: 60/212,681
/ PRIOR FILING DATE: 2000-06-20
/ PRIOR APPLICATION NUMBER: 60/290,170
/ PRIOR FILING DATE: 2001-05-10
/ PRIOR APPLICATION NUMBER: PA 2001 00739
/ PRIOR FILING DATE: 2001-05-10
/ NUMBER OF SEQ ID NOS: 24
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 239
/ TYPE: PRT
/ ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match
Best Local Similarity 99.2%; Score 1264; DB 12; Length 239;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHHYLSQSALSQKDPNEKRDHVMVLCFVTAAGITTLGMDLEYK 239
DB 181 DHYQONTPIGDGPVLLPDNHHYLSQSALSQKDPNEKRDHVMVLCFVTAAGITTLGMDLEYK 239

RESULT 8
US-09-920-922-2
/ Sequence 2, Application US/09920922
/ Patent No. US20020083488A1
/ GENERAL INFORMATION:
/ APPLICANT: Miyawaki, Atsushi
/ APPLICANT: Sawano, Asako
/ TITLE OF INVENTION: METHOD FOR MUTAGENESIS
/ FILE REFERENCE: 11283-012001
/ CURRENT APPLICATION NUMBER: US/09/920,922
/ CURRENT FILING DATE: 2001-08-02
/ PRIOR APPLICATION NUMBER: JP 2000-237166
/ PRIOR FILING DATE: 2000-08-04
/ NUMBER OF SEQ ID NOS: 9
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 2
/ LENGTH: 239
/ TYPE: PRT
/ ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match
Best Local Similarity 99.0%; Score 1261; DB 9; Length 239;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGKLTTLKFICTTGKLPVPWPT 60
```

```

Qy 61 LVTTLSYGVQCFSPYDPHMKQHDFFKSAPEGYVOERTIFFKDDGNYKTRAEVKPEGDTL 120
Db 61 LVTTLTYGVCFSRYPPDHMKQHDFFKSAPEGYVOERTIFFKDDGNYKTRAEVKPEGDTL 120
Qy 121 VNRIELKGIDFKEDGNIIGHKL EYNNYNHNHVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNIIGHKL EYNNYNHNHVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Qy 181 DHYQOQNTPIGDPGVLLPDNHYLSTOSALS KDPNEKRDRHMVLFGFVTAAGITLGMDELYK 239
Db 181 DHYQOQNTPIGDPGVLLPDNHYLSTOSALS KDPNEKRDRHMVLFGFVTAAGITLGMDELYK 239

RESULT 9
US-09-999-745-4
; Sequence 4, Application US/09999745
; Patent No. US20020157120A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGEN1470-1
; CURRENT APPLICATION NUMBER: US/09/999,745
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
; PRIOR FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: patent in version 3.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-999-745-4

```

Query Match	99.0%;	Score 1261;	DB 9;	Length 239;
Best Local Similarity	98.7%;	Pred. No. 3.1e-123;		
Matches 236;	Conservative 2;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	1	MVSGEELFTGVVPIILVELGDVNGHKFVS	GEGBDATY	GKLTILKFICTTGKLPVPWPT 60
Db	1	MVSGEELFTGVVPIILVELGDVNGHKFVS	GEGBDATY	GKLTILKFICTTGKLPVPWPT 60
Qy	51	LVTTLSYGVQCFSRYPDHMKQHDFFK	SAMPEGYVQERTIFFKDGNYKTRAEV	FEGBDTL 120
Db	51	LVTTLTYGVCQFSRYPDHMKQHDFFK	SAMPEGYVQERTIFFKDGNYKTRAEV	FEGBDTL 120
Qy	121	VNRIELKGIQDFKEDGNILGHKLEFYN	YNSHNVIWADQKQNGIKYNEKIRHNI	EDGSVOLA 180
Db	121	VNRIELKGIQDFKEDGNILGHKLEFYN	YNSHNVIWADQKQNGIKYNEKIRHNI	EDGSVOLA 180
Qy	131	DHYQONTPIGDGPVLLPDNHYLSTQS	ALS	KDNPNEKRDHMLVIGFTVAAGITLGMDELYK 239
Db	131	DHYQONTPIGDGPVLLPDNHYLSTQS	ALS	KDNPNEKRDHMLVIGFTVAAGITLGMDELYK 239

```

RESULT 10
US-09-866-538-4
; Sequence 4, Application US/09866538
; Publication No. US2003032088A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TS'EN, ROGER
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/09/866,538
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT

```

```

; ORGANISM: Aequorea victoria
US-09-866-538-4

Query Match          99.0%:   Score 1261;   DB 10;   Length 239;
Best Local Similarity 98.7%:   Pred. No. 3.1e-123;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MVSKEELFTGVPILVELDGDVNGHKFVSVEGEGDATYGLKTLKFICTTGKLPVPWPT 60
Db 1 MVSKEELFTGVPILVELDGDVNGHKFVSVEGEGDATYGLKTLKFICTTGKLPVPWPT 60

Qy 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIEFKDDGNYKTRAEVKFEGDTL 120
Db 61 LVTTLTATGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIEFKDDGNYKTRAEVKFEGDTL 120

Qy 121 VNRIELKGIDFKEDGNLGHLEKLYNNVNSHNVYIMADKQKNGIKVNFKIRNIEDGSVQLA 180
Db 121 VNRIELKGIDFKEDGNLGHLEKLYNNVNSHNVYIMADKQKNGIKVNFKIRNIEDGSVQLA 180

Qy 181 DHYQONTPIGDGPVLLPDNHYLTSQSALSXPDPNEKRDHVMVLIGFVTAAGITLGMDELYK 239
Db 181 DHYQONTPIGDGPVLLPDNHYLTSQSALSXPDPNEKRDHVMVLLEFVTAAGITLGMDELYK 239

RESULT 11
US-09-797-496B-2
; Sequence 2, Application US/09797496B
; Publication No. US20030049597A1
; GENERAL INFORMATION:
; APPLICANT: Simon, Sanford M.
; APPLICANT: Chen, Yu
; TITLE OF INVENTION: Chimeric Fluorescent Enzymes and Uses Thereof
; FILE REFERENCE: 600-1-267
; CURRENT APPLICATION NUMBER: US/09/797,496B
; CURRENT FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Aequorea victoria green fluorescent protein modified as
US-09-797-496B-2
; OTHER INFORMATION: in specification
US-09-797-496B-2

```

Query Match	99.0%	Score 1261;	DB 10;	Length 239;
Best Local Similarity	98.7%;	Pred. No. 3.1e-123;		
Matches 236;	Conservative 2;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	1	MVSKGRELFTGVVPIILVELDGDVNGHKFVSGBEGDATYVKLTTLFICTTGKLPVPWPT	60	
Db	1	MVSKGRELFTGVVPIILVELDGDVNGHKFVSGBEGDATYVKLTTLFICTTGKLPVPWPT	60	
Qy	61	LVTTLSYGVQCFRYPDHHKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL	120	
Db	61	LVTTLSYGVQCFRYPDHHKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL	120	
Qy	121	VNRIELKGDIDFKEDGNILGHLELYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA	180	
Db	121	VNRIELKGDIDFKEDGNILGHLELYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA	180	
Qy	181	DHYQONTPIGDGPVLLPDNHYLSTQSALSKOPNEKSDHVMVLIGFVTAAGITLGMDELYK	239	
Db	181	DHYQONTPIGDGPVLLPDNHYLSTQSALSKOPNEKSDHVMVLIGFVTAAGITLGMDELYK	239	

RESULT 12  
US-09-794-308--4  
; Sequence 4, Application US/09794308  
; Publication No. US20030170911A1  
; GENERAL INFORMATION:  
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA

APPLICANT: TSUEN, Roger  
APPLICANT: ZACHARIAS, David  
APPLICANT: BAIRD, Geoffrey  
TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS  
FILE REFERENCE: REGEN1530  
CURRENT APPLICATION NUMBER: US/09/794,308  
CURRENT FILING DATE: 2001-02-26  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
US-09-794-308-4

Query Match 99.0%; Score 1261; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 3.1e-123;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGLTLKFKICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGLTLKFKICTTGKLPVPWPT 60  
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSQKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSQKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239

RESULT 13  
US-09-865-291-4  
Sequence 4, Application US/09865291  
Publication No. US20030186229A1  
GENERAL INFORMATION:  
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA  
APPLICANT: TSUEN, Roger  
APPLICANT: TING, Alice  
APPLICANT: ZHANG, Jin  
TITLE OF INVENTION: EMISSION RATIONETRIC INDICATORS OF PHOSPHORYLATION  
FILE REFERENCE: REGEN1550  
CURRENT APPLICATION NUMBER: US/09/865,291  
CURRENT FILING DATE: 2001-05-24  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
US-09-865-291-4

Query Match 99.0%; Score 1261; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 3.1e-123;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGLTLKFKICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGLTLKFKICTTGKLPVPWPT 60  
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSQKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSQKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239  
RESULT 14  
US-10-457-982-3  
Sequence 3, Application US/10457982  
Publication No. US20030212265A1  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger Y.  
APPLICANT: Miyawaki, Acsuehi  
APPLICANT: Lloplis, Juan  
APPLICANT: Wächter, Rebekka M.  
APPLICANT: Remington, S. James  
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
FILE REFERENCE: 07257/071001  
CURRENT APPLICATION NUMBER: US/10/457,982  
CURRENT FILING DATE: 2003-06-09  
PRIOR APPLICATION NUMBER: US/09/602,641  
PRIOR FILING DATE: 2000-06-22  
PRIOR APPLICATION NUMBER: 09/172,063  
PRIOR FILING DATE: 1998-10-13  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (0)...(0)  
OTHER INFORMATION: EGFP  
US-10-457-982-3

Query Match 99.0%; Score 1261; DB 12; Length 239;  
Best Local Similarity 98.7%; Pred. No. 3.1e-123;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGLTLKFKICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGLTLKFKICTTGKLPVPWPT 60  
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSQKDPNEKRDHMLVIGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNNHYLSTQSALSQKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239

RESULT 15  
US-10-121-258-13  
Sequence 13, Application US/10121258  
Publication No. US20030059835A1  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger  
APPLICANT: Campbell, Robert  
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT  
TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME  
FILE REFERENCE: UC083.1CP2CP1  
CURRENT APPLICATION NUMBER: US/10/121,258  
CURRENT FILING DATE: 2002-04-10  
PRIOR APPLICATION NUMBER: 09/794,308  
PRIOR FILING DATE: 2001-02-26  
PRIOR APPLICATION NUMBER: 09/866,538  
PRIOR FILING DATE: 2001-05-24

```
; NUMBER CF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NC 13
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Enhanced Green Fluorescent Protein (EGFP)
US-10-121-258-13

Query Match          99.0%; Score 1261; DB 14; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.1e-123;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 MVSKEELFTGVVPILVELDGDVNGHKFSPVSGEGDGYGKLTILKFICTTGKLPVPWPT 60
Db      1 MVSKEELFTGVVPILVELDGDVNGHKFSPVSGEGDGYGKLTILKFICTTGKLPVPWPT 60

QY      61 LVTTLSYGVOCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db      61 LVTTLTYGVOCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY      121 VNRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
Db      121 VNRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY      181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHWLVLEFVTAAGITLGMDELYK 239
Db      181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHWLVLEFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:09:28
Job time : 34.7778 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:00 ; Search time 10.3333 Seconds  
(without alignments)  
2224.817 Million cell updates/sec

Title: US-09-887-784-222i

Perfect score: 1274

Sequence: 1 MVSKEELFTGVVILVELD.....VLIGFVTRAGITLGMDELYK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1239	97.3	238	1 JQ1514	green-fluorescent
2	104	8.2	785	2 H72228	hypothetical prote
3	91.5	7.2	861	2 H64102	leucine-trna ligas
4	90	7.1	632	2 T06586	DNA-binding protei
5	89.5	7.0	887	2 E82590	leucyl-trna synth
6	88.5	6.9	655	2 D83917	DNA topoisomerase
7	88.5	6.9	797	2 JC4078	protective surface
8	88.5	6.9	808	2 F64102	protective surface
9	87.5	6.9	370	2 E70390	iron-sulfur cofact
10	87.5	6.9	860	2 AC0582	leucyl-trna synth
11	87.5	6.9	941	2 S29043	cellulase (EC 3.2.
12	87.5	6.9	2222	1 A36028	DNA-directed DNA p
13	87.5	6.9	2573	2 D71614	hypothetical prote
14	87	6.8	578	1 I40794	dihydrolipoamide d
15	86	6.8	357	2 G81355	tRNA (uracil-5-) m
16	85.5	6.7	788	1 JDLV1H	DNA-directed DNA p
17	85.5	6.7	889	2 JC5576	inter-alpha-trypsi
18	85	6.7	281	2 AD2052	hypothetical prote
19	85	6.7	874	2 JC4930	S-layer protein pr
20	84.5	6.6	425	2 C97354	hypothetical prote
21	84.5	6.6	613	2 A99552	oligoendopeptidase
22	84	6.6	353	2 E84941	imidazoleglycerol-
23	84	6.6	461	2 T06936	photosystem II chl
24	83.5	6.6	836	1 JDLVD	DNA-directed DNA p
25	83.5	6.6	1134	2 A60234	Iga Fc receptor pr
26	83.5	6.6	1164	1 FCSOAG	Iga Fc receptor pr
27	83	6.6	439	2 JH0414	synaptogamin o-p65
28	82.5	6.5	534	1 NICLMA	nitrogenase (EC 1.
29	82.5	6.5	740	2 G95153	neuraminidase, pro

RESULT 1

JQ1514  
green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)  
C:Species: Aequorea victoria  
C>Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 23-Mar-2001  
C:Accession: JQ1514; PQ0335; S48693; S51330; S51331  
R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.  
Gene ill, 229-233, 1992  
A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.  
A:Reference number: JQ1514; MUID:92175527; PMID:1347277  
A:Accession: JQ1514  
A:Molecule type: DNA  
A:Residues: 1-107, 'S', 109-238 <PRA1>  
A:Cross-references: GB:M62654; NID:G155662; PIDN:AAA27722.1; PID:G155663  
A:Accession: JQ1514  
A:Molecule type: mRNA  
A:Residues: 1-99, 'F', 101-140, 'L', 142-218, 'V', 220-238 <PRA2>  
A:Cross-references: GB:M62653; NID:G155660; PIDN:AAA27721.1; PID:G155661  
A:Accession: PQ0335  
A:Molecule type: protein  
A:Residues: 46-64; 74-122; 132-151; 154-183; 185-200 <PRA3>  
R:Inouye, S.; Tsuji, F.I.  
FEBS Lett. 351, 211-214, 1994  
A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.  
A:Reference number: S48693; MUID:94364470; PMID:8082767  
A:Accession: S48693  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-24, 'O', 26-156, 'P', 158-171, 'K', 173-238 <INO>  
A:Cross-references: GB:L29345; NID:G606383; PIDN:AAA58246.1; PID:G606384  
R:Watkins, J.N.; Campbell, A.K.  
submitted to the EMBL Data Library, January 1995  
A:Reference number: S51330  
A:Accession: S51330  
A:Molecule type: mRNA  
A:Residues: 1-13, 'V', 15-24, 'O', 26-44, 'N', 46-153, 'G', 155-156, 'P', 158-171, 'K', 173-227, 'R', 229-238 <PRA4>  
A:Cross-references: EMBL:X83959; NID:G634008; PIDN:CAA58789.1; PID:G634009  
A:Experimental source: clone gfp1  
A:Accession: S51331  
A:Molecule type: mRNA  
A:Residues: 1-24, 'O', 26-29, 'R', 31-83, 'L', 85-153, 'G', 155-156, 'P', 158-171, 'K', 173-208, 'Q', 209-238 <PRA5>  
A:Cross-references: EMBL:X83960; NID:G634010; PIDN:CAA58790.1; PID:G634011  
A:Experimental source: clone gfp2  
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.  
submitted to the Brookhaven Protein Data Bank, August 1996  
A:Reference number: A65692; PDB:1GFL  
A:Reference: annotation: X-ray crystallography, 1.9 angstroms, residues 'A', 2-79, 'R', 81-95, 'L', 97-100, 'K', 102-103, 'G', 105-106, 'D', 108-109, 'E', 111-112, 'K', 114-115, 'G', 117-118, 'K', 120-121, 'G', 123-124, 'K', 126-127, 'G', 129-130, 'K', 132-133, 'G', 135-136, 'K', 138-139, 'G', 141-142, 'K', 144-145, 'G', 147-148, 'K', 150-151, 'G', 153-154, 'K', 156-157, 'G', 159-160, 'K', 162-163, 'G', 165-166, 'K', 168-169, 'G', 171-172, 'K', 174-175, 'G', 177-178, 'K', 180-181, 'G', 183-184, 'K', 186-187, 'G', 189-190, 'K', 192-193, 'G', 195-196, 'K', 198-199, 'G', 201-202, 'K', 204-205, 'G', 207-208, 'K', 210-211, 'G', 213-214, 'K', 216-217, 'G', 219-220, 'K', 222-223, 'G', 225-226, 'K', 228-229, 'G', 231-232, 'K', 234-235, 'G', 237-238, 'K', 240-241, 'G', 243-244, 'K', 246-247, 'G', 249-250, 'K', 252-253, 'G', 255-256, 'K', 258-259, 'G', 261-262, 'K', 264-265, 'G', 267-268, 'K', 270-271, 'G', 273-274, 'K', 276-277, 'G', 279-280, 'K', 282-283, 'G', 285-286, 'K', 288-289, 'G', 291-292, 'K', 294-295, 'G', 297-298, 'K', 300-301, 'G', 303-304, 'K', 306-307, 'G', 309-310, 'K', 312-313, 'G', 315-316, 'K', 318-319, 'G', 321-322, 'K', 324-325, 'G', 327-328, 'K', 330-331, 'G', 333-334, 'K', 336-337, 'G', 339-340, 'K', 342-343, 'G', 345-346, 'K', 348-349, 'G', 351-352, 'K', 354-355, 'G', 357-358, 'K', 360-361, 'G', 363-364, 'K', 366-367, 'G', 369-370, 'K', 372-373, 'G', 375-376, 'K', 378-379, 'G', 381-382, 'K', 384-385, 'G', 387-388, 'K', 390-391, 'G', 393-394, 'K', 396-397, 'G', 399-400, 'K', 402-403, 'G', 405-406, 'K', 408-409, 'G', 411-412, 'K', 414-415, 'G', 417-418, 'K', 420-421, 'G', 423-424, 'K', 426-427, 'G', 429-430, 'K', 432-433, 'G', 435-436, 'K', 438-439, 'G', 441-442, 'K', 444-445, 'G', 447-448, 'K', 450-451, 'G', 453-454, 'K', 456-457, 'G', 459-460, 'K', 462-463, 'G', 465-466, 'K', 468-469, 'G', 471-472, 'K', 474-475, 'G', 477-478, 'K', 480-481, 'G', 483-484, 'K', 486-487, 'G', 489-490, 'K', 492-493, 'G', 495-496, 'K', 498-499, 'G', 501-502, 'K', 504-505, 'G', 507-508, 'K', 510-511, 'G', 513-514, 'K', 516-517, 'G', 519-520, 'K', 522-523, 'G', 525-526, 'K', 528-529, 'G', 531-532, 'K', 534-535, 'G', 537-538, 'K', 540-541, 'G', 543-544, 'K', 546-547, 'G', 549-550, 'K', 552-553, 'G', 555-556, 'K', 558-559, 'G', 561-562, 'K', 564-565, 'G', 567-568, 'K', 570-571, 'G', 573-574, 'K', 576-577, 'G', 579-580, 'K', 582-583, 'G', 585-586, 'K', 588-589, 'G', 591-592, 'K', 594-595, 'G', 597-598, 'K', 600-601, 'G', 603-604, 'K', 606-607, 'G', 609-610, 'K', 612-613, 'G', 615-616, 'K', 618-619, 'G', 621-622, 'K', 624-625, 'G', 627-628, 'K', 630-631, 'G', 633-634, 'K', 636-637, 'G', 639-640, 'K', 642-643, 'G', 645-646, 'K', 648-649, 'G', 651-652, 'K', 654-655, 'G', 657-658, 'K', 660-661, 'G', 663-664, 'K', 666-667, 'G', 669-670, 'K', 672-673, 'G', 675-676, 'K', 678-679, 'G', 681-682, 'K', 684-685, 'G', 687-688, 'K', 690-691, 'G', 693-694, 'K', 696-697, 'G', 699-700, 'K', 702-703, 'G', 705-706, 'K', 708-709, 'G', 711-712, 'K', 714-715, 'G', 717-718, 'K', 720-721, 'G', 723-724, 'K', 726-727, 'G', 729-730, 'K', 732-733, 'G', 735-736, 'K', 738-739, 'G', 741-742, 'K', 744-745, 'G', 747-748, 'K', 750-751, 'G', 753-754, 'K', 756-757, 'G', 759-760, 'K', 762-763, 'G', 765-766, 'K', 768-769, 'G', 771-772, 'K', 774-775, 'G', 777-778, 'K', 780-781, 'G', 783-784, 'K', 786-787, 'G', 789-790, 'K', 792-793, 'G', 795-796, 'K', 798-799, 'G', 801-802, 'K', 804-805, 'G', 807-808, 'K', 810-811, 'G', 813-814, 'K', 816-817, 'G', 819-820, 'K', 822-823, 'G', 825-826, 'K', 828-829, 'G', 831-832, 'K', 834-835, 'G', 837-838, 'K', 840-841, 'G', 843-844, 'K', 846-847, 'G', 849-850, 'K', 852-853, 'G', 855-856, 'K', 858-859, 'G', 861-862, 'K', 864-865, 'G', 867-868, 'K', 870-871, 'G', 873-874, 'K', 876-877, 'G', 879-880, 'K', 882-883, 'G', 885-886, 'K', 888-889, 'G', 891-892, 'K', 894-895, 'G', 897-898, 'K', 900-901, 'G', 903-904, 'K', 906-907, 'G', 909-910, 'K', 912-913, 'G', 915-916, 'K', 918-919, 'G', 921-922, 'K', 924-925, 'G', 927-928, 'K', 930-931, 'G', 933-934, 'K', 936-937, 'G', 939-940, 'K', 942-943, 'G', 945-946, 'K', 948-949, 'G', 951-952, 'K', 954-955, 'G', 957-958, 'K', 960-961, 'G', 963-964, 'K', 966-967, 'G', 969-970, 'K', 972-973, 'G', 975-976, 'K', 978-979, 'G', 981-982, 'K', 984-985, 'G', 987-988, 'K', 990-991, 'G', 993-994, 'K', 996-997, 'G', 999-1000, 'K', 1002-1003, 'G', 1005-1006, 'K', 1008-1009, 'G', 1011-1012, 'K', 1014-1015, 'G', 1017-1018, 'K', 1020-1021, 'G', 1023-1024, 'K', 1026-1027, 'G', 1029-1030, 'K', 1032-1033, 'G', 1035-1036, 'K', 1038-1039, 'G', 1041-1042, 'K', 1044-1045, 'G', 1047-1048, 'K', 1050-1051, 'G', 1053-1054, 'K', 1056-1057, 'G', 1059-1060, 'K', 1062-1063, 'G', 1065-1066, 'K', 1068-1069, 'G', 1071-1072, 'K', 1074-1075, 'G', 1077-1078, 'K', 1080-1081, 'G', 1083-1084, 'K', 1086-1087, 'G', 1089-1090, 'K', 1092-1093, 'G', 1095-1096, 'K', 1098-1099, 'G', 1101-1102, 'K', 1104-1105, 'G', 1107-1108, 'K', 1110-1111, 'G', 1113-1114, 'K', 1116-1117, 'G', 1119-1120, 'K', 1122-1123, 'G', 1125-1126, 'K', 1128-1129, 'G', 1131-1132, 'K', 1134-1135, 'G', 1137-1138, 'K', 1140-1141, 'G', 1143-1144, 'K', 1146-1147, 'G', 1149-1150, 'K', 1152-1153, 'G', 1155-1156, 'K', 1158-1159, 'G', 1161-1162, 'K', 1164-1165, 'G', 1167-1168, 'K', 1170-1171, 'G', 1173-1174, 'K', 1176-1177, 'G', 1179-1180, 'K', 1182-1183, 'G', 1185-1186, 'K', 1188-1189, 'G', 1191-1192, 'K', 1194-1195, 'G', 1197-1198, 'K', 1200-1201, 'G', 1203-1204, 'K', 1206-1207, 'G', 1209-1210, 'K', 1212-1213, 'G', 1215-1216, 'K', 1218-1219, 'G', 1221-1222, 'K', 1224-1225, 'G', 1227-1228, 'K', 1230-1231, 'G', 1233-1234, 'K', 1236-1237, 'G', 1239-1240, 'K', 1242-1243, 'G', 1245-1246, 'K', 1248-1249, 'G', 1251-1252, 'K', 1254-1255, 'G', 1257-1258, 'K', 1260-1261, 'G', 1263-1264, 'K', 1266-1267, 'G', 1269-1270, 'K', 1272-1273, 'G', 1275-1276, 'K', 1278-1279, 'G', 1281-1282, 'K', 1284-1285, 'G', 1287-1288, 'K', 1290-1291, 'G', 1293-1294, 'K', 1296-1297, 'G', 1299-1300, 'K', 1302-1303, 'G', 1305-1306, 'K', 1308-1309, 'G', 1311-1312, 'K', 1314-1315, 'G', 1317-1318, 'K', 1320-1321, 'G', 1323-1324, 'K', 1326-1327, 'G', 1329-1330, 'K', 1332-1333, 'G', 1335-1336, 'K', 1338-1339, 'G', 1341-1342, 'K', 1344-1345, 'G', 1347-1348, 'K', 1350-1351, 'G', 1353-1354, 'K', 1356-1357, 'G', 1359-1360, 'K', 1362-1363, 'G', 1365-1366, 'K', 1368-1369, 'G', 1371-1372, 'K', 1374-1375, 'G', 1377-1378, 'K', 1380-1381, 'G', 1383-1384, 'K', 1386-1387, 'G', 1389-1390, 'K', 1392-1393, 'G', 1395-1396, 'K', 1398-1399, 'G', 1401-1402, 'K', 1404-1405, 'G', 1407-1408, 'K', 1410-1411, 'G', 1413-1414, 'K', 1416-1417, 'G', 1419-1420, 'K', 1422-1423, 'G', 1425-1426, 'K', 1428-1429, 'G', 1431-1432, 'K', 1434-1435, 'G', 1437-1438, 'K', 1440-1441, 'G', 1443-1444, 'K', 1446-1447, 'G', 1449-1450, 'K', 1452-1453, 'G', 1455-1456, 'K', 1458-1459, 'G', 1461-1462, 'K', 1464-1465, 'G', 1467-1468, 'K', 1470-1471, 'G', 1473-1474, 'K', 1476-1477, 'G', 1479-1480, 'K', 1482-1483, 'G', 1485-1486, 'K', 1488-1489, 'G', 1491-1492, 'K', 1494-1495, 'G', 1497-1498, 'K', 1500-1501, 'G', 1503-1504, 'K', 1506-1507, 'G', 1509-1510, 'K', 1512-1513, 'G', 1515-1516, 'K', 1518-1519, 'G', 1521-1522, 'K', 1524-1525, 'G', 1527-1528, 'K', 1530-1531, 'G', 1533-1534, 'K', 1536-1537, 'G', 1539-1540, 'K', 1542-1543, 'G', 1545-1546, 'K', 1548-1549, 'G', 1551-1552, 'K', 1554-1555, 'G', 1557-1558, 'K', 1560-1561, 'G', 1563-1564, 'K', 1566-1567, 'G', 1569-1570, 'K', 1572-1573, 'G', 1575-1576, 'K', 1578-1579, 'G', 1581-1582, 'K', 1584-1585, 'G', 1587-1588, 'K', 1590-1591, 'G', 1593-1594, 'K', 1596-1597, 'G', 1599-1600, 'K', 1602-1603, 'G', 1605-1606, 'K', 1608-1609, 'G', 1611-1612, 'K', 1614-1615, 'G', 1617-1618, 'K', 1620-1621, 'G', 1623-1624, 'K', 1626-1627, 'G', 1629-1630, 'K', 1632-1633, 'G', 1635-1636, 'K', 1638-1639, 'G', 1641-1642, 'K', 1644-1645, 'G', 1647-1648, 'K', 1650-1651, 'G', 1653-1654, 'K', 1656-1657, 'G', 1659-1660, 'K', 1662-1663, 'G', 1665-1666, 'K', 1668-1669, 'G', 1671-1672, 'K', 1674-1675, 'G', 1677-1678, 'K', 1680-1681, 'G', 1683-1684, 'K', 1686-1687, 'G', 1689-1690, 'K', 1692-1693, 'G', 1695-1696, 'K', 1698-1699, 'G', 1701-1702, 'K', 1704-1705, 'G', 1707-1708, 'K', 1710-1711, 'G', 1713-1714, 'K', 1716-1717, 'G', 1719-1720, 'K', 1722-1723, 'G', 1725-1726, 'K', 1728-1729, 'G', 1731-1732, 'K', 1734-1735, 'G', 1737-1738, 'K', 1740-1741, 'G', 1743-1744, 'K', 1746-1747, 'G', 1749-1750, 'K', 1752-1753, 'G', 1755-1756, 'K', 1758-1759, 'G', 1761-1762, 'K', 1764-1765, 'G', 1767-1768, 'K', 1770-1771, 'G', 1773-1774, 'K', 1776-1777, 'G', 1779-1780, 'K', 1782-1783, 'G', 1785-1786, 'K', 1788-1789, 'G', 1791-1792, 'K', 1794-1795, 'G', 1797-1798, 'K', 1800-1801, 'G', 1803-1804, 'K', 1806-1807, 'G', 1809-1810, 'K', 1812-1813, 'G', 1815-1816, 'K', 1818-1819, 'G', 1821-1822, 'K', 1824-1825, 'G', 1827-1828, 'K', 1830-1831, 'G', 1833-1834, 'K', 1836-1837, 'G', 1839-1840, 'K', 1842-1843, 'G', 1845-1846, 'K', 1848-1849, 'G', 1851-1852, 'K', 1854-1855, 'G', 1857-1858, 'K', 1860-1861, 'G', 1863-1864, 'K', 1866-1867, 'G', 1869-1870, 'K', 1872-1873, 'G', 1875-1876, 'K', 1878-1879, 'G', 1881-1882, 'K', 1884-1885, 'G', 1887-1888, 'K', 1890-1891, 'G', 1893-1894, 'K', 1896-1897, 'G', 1899-1900, 'K', 1902-1903, 'G', 1905-1906, 'K', 1908-1909, 'G', 1911-1912, 'K', 1914-1915, 'G', 1917-1918, 'K', 1920-1921, 'G', 1923-1924, 'K', 1926-1927, 'G', 1929-1930, 'K', 1932-1933, 'G', 1935-1936, 'K', 1938-1939, 'G', 1941-1942, 'K', 1944-1945, 'G', 1947-1948, 'K', 1950-1951, 'G', 1953-1954, 'K', 1956-1957, 'G', 1959-1960, 'K', 1962-1963, 'G', 1965-1966, 'K', 1968-1969, 'G', 1971-1972, 'K', 1974-1975, 'G', 1977-1978, 'K', 1980-1981, 'G', 1983-1984, 'K', 1986-1987, 'G', 1989-1990, 'K', 1992-1993, 'G', 1995-1996, 'K', 1998-1999, 'G', 2001-2002, 'K', 2004-2005, 'G', 2007-2008, 'K', 2010-2011, 'G', 2013-2014, 'K', 2016-2017, 'G', 201

A;Contents: annotation; X-ray crystallography, 1.9 angstroms  
C;Comment: This protein is excited by the photoprotein sequeirin (see PIR:AQJFNV) emitting  
C;Comment: The chromophore of this protein is formed by modification of Ser-dehydro-Tyr-  
C;Genetics:  
A;Gene: GFP  
A;Introns: 69/3; 167/3  
A;Superfamily: green-fluorescent protein  
C;Keywords: chromoprotein; luminescence  
F;55-67/Cross-Link: 5-Imidazolinone (Ser-Gly) #status experimental  
F;66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match 97.3%; Score 1239; DB 1; Length 238;  
Best Local Similarity 96.6%; Pred. No. 3.2e-96;  
Matches 230; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGLTKLTKFICTTGKLPVWPPTL 61  
Db 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGLTKLTKFICTTGKLPVWPPTL 60

QY 62 VTTLTGVOCFSRYPDHMKQHDFFKSAMPEGYVOERTIPFKDDGNYKTRAEVKFEGETLV 121  
Db 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGYVOERTIPFKDDGNYKTRAEVKFEGETLV 120

QY 122 NRLELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181  
Db 121 NRLELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQSALSKDPNEKRDMVLIGFVTAAGITLGMDELYK 239  
Db 181 HYQONTPIGDGPVLLPDNHYLSQSALSKDPNEKRDMVLIGFVTAAGITLGMDELYK 238

RESULT 2  
H72228  
hypothetical protein TM1624 - Thermotoga maritima (strain MSB8)  
C;Species: Thermotoga maritima  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C;Accession: H72228  
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.

Nature 399, 323-329, 1999  
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A;Reference number: A72200; MUID:99287316; PMID:10360571  
A;Accession: H72228  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-785 <ARN>  
A;Cross-references: GB:AE001806; GB:AE000512; NID:g4982196; PIDN:AAD36691.1; PID:g498219  
A;Experimental source: strain MSB8  
C;Genetics:  
A;Gene: TM1624

Query Match 8.2%; Score 104; DB 2; Length 785;  
Best Local Similarity 19.7%; Pred. No. 1.1;  
Matches 46; Conservative 32; Mismatches 71; Indels 84; Gaps 7;

QY 3 SKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGLTKLTKFICTTGKLPVWPPTLV 62  
Db 15 NEGRFSFEGTVPGVQAD-----LVRKGLLPHYPVGM- 46

QY 63 TTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVOERTIPFKDDGNYKTRAEVKFEGETLVN 122  
Db 47 -----NEDLFKEIDREWIYERFEFKEDVKGERVDLVFEGVDTLVS 88

QY 123 RIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFKIRHNIEDGSVOLADH 182  
Db 89 DVTLVNGVYL---GSTEDMFEYFDFTNVL-----KEKNHLKVYIK-----SPIRVKPT 134

QY 183 YQONTPIGDGPVLLPDNHYLSQSALSKDPNEKRDMVLIGFVTAAGITLGM 235  
Db 135 LEQNYGLVGGP-----EDP-----IRGYIRKAQYSYGM 163

RESULT 3

H64102 leucine-tRNA ligase (EC 6.1.1.4) - Haemophilus influenzae (strain Rd KW20)

N;Alternate names: leucyl-tRNA synthetase

C;Species: Haemophilus influenzae

C;Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 03-Jun-2002

C;Accession: H64102

R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.  
; Gocayne, J.D.; Scott, J.; Shifley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995

A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, C.

A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A;Reference number: A64000; MUID:95350630; PMID:7542800

A;Accession: H64102

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-861 <TIGR>

C;Genetics:

A;Gene: leuS

C;Superfamily: leucine-tRNA ligase

C;Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 7.2%; Score 91.5; DB 2; Length 861;  
Best Local Similarity 24.1%; Pred. No. 14;  
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;

QY 50 TTGKLPVPWPPTLVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVOERTIPFKD----- 103  
Db 314 TGDKLPI-WVANFVLMHYGTGAVMAVPAH-DQDQDF-----EFAQKYSLPKQVIAPLA 364

QY 104 DGNKYTRAEVKFEGETLVNRIELKGIDFKEDGNILGHKLEYNNSHVYIMADK-QKNGI 162  
Db 365 DEEIDLTKQAFVEHGHKLVNSDFDGKNF--DGAFNG-----IADKLEKLG 408

QY 163 ---KYNFKIRH-----NIEDGSVOLADHYQQWTPIGDGPVLLPDNHYL- 202  
Db 409 GKRQVNIIRDNGVSRQRYWGAPIPMLTLENGDVPA-----PMEDLPILLPEDVMD 461

QY 203 STQSALSKDPN 213  
Db 462 GVKSPINADPN 472

RESULT 4

T06586

DNA-binding protein PD2 - garden pea

C;Species: Pisum sativum (garden pea)

C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 08-Oct-1999

C;Accession: T06586

R;Sato, N.; Kazuno, A.A.; Ohta, N.; Ohshima, K.

submitted to the EMBL Data Library, June 1996

A;Description: Identification of a novel family of DNA-binding proteins with two AT-hook

A;Reference number: Z15774

A;Accession: T06586

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-632 <SAT>

A;Cross-references: EMBL:X98740; NID:e995229; PIDN:CAA67292.1; PID:e275185

A;Experimental source: cv. Alaska

Query Match 7.1%; Score 90; DB 2; Length 632;  
Best Local Similarity 23.3%; Pred. No. 12;  
Matches 49; Conservative 26; Mismatches 79; Indels 56; Gaps 7;

QY 16 LVELDGDVNGHKFSVSGEGDATYGLTKLTKFICTTGKLPVWPPTLVTTLSYGVQCFSR 75  
Db 363 IVDQGRDVGSKVDVINKESNEATIPENK-----PTEPKLDVQEQLAATM----- 408

QY 76 PDHMKQHDFFKSAMPEGYV-----QERTIFFKDDGNYKTRAEVKFEGETLVNRIE 125



[illegible]

Db 543 RNLIYQSMKFKNGIKTN-----DFDFSGFWNNSLNLRGYPFKGVKASLG-GRVTI 593  
QY 197 P--DNHYLSTQSALSADPNKRDHMLVIGFVTAAGITLG 233  
Db 594 PGSDNKYYKLSADVQGFYPLDRDHLWVVSASAGAGYANG 632  
RESULT 8  
F64102  
protective surface antigen D-15 - Haemophilus influenzae (strain Rd KW20)  
C:Species: Haemophilus influenzae  
C>Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 18-Sep-1998  
C:Accession: F64102  
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J. D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M. Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630; PMID:7542800  
A:Accession: F64102  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-808 <TIGR>  
A:Cross-references: GB:L42023; TIGR:HI0917  
C:Superfamily: protective surface antigen D-15  
C:Keywords: surface antigen  
Query Match 6.9%; Score 88.5; DB 2; Length 808;  
Best Local Similarity 21.9%; Pred. No. 23;  
Matches 48; Conservative 30; Mismatches 78; Indels 63; Gaps 11;  
QY 65 LSVGVQCFSRYPDHMKQHDF-----FKSAMPEGYVQOE-----RTI 99  
Db 440 IGYGTESGISYQASVKQDNFLGTGCAAVSIAGTKNDYGTSVNLGYTEPYFTKDGVSILGNNV 499  
QY 100 FFKDDGNYKTRAEVKEGDTLVNRIELKIGDFKEDGNI---LGH-----KLEYNYS 148  
Db 500 FFENYDNKSDTSNSNKRITTYGNSVTL-GFPVNNNSYVGLGHTYNTKSNFALEYN--- 555  
QY 149 HNYVIMADKQK-NGIKVNFKIRHNIEDSGSVQLADHYQQ-----NTPIGDGPVLL 196  
Db 556 RNLIYQSMKFKNGIKTN-----DFDFSGFWNNSLNLRGYPFKGVKASLG-GRVTI 606  
QY 197 P--DNHYLSTQSALSADPNKRDHMLVIGFVTAAGITLG 233  
Db 607 PGSDNKYYKLSADVQGFYPLDRDHLWVVSASAGAGYANG 645  
RESULT 9  
E70390  
iron-sulfur cofactor synthesis protein nifs - Aquifex aeolicus  
N:Contains: L-cysteine sulfoxylase (EC 2.8.1.-)  
C:Species: Aquifex aeolicus  
C>Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 07-Dec-1999  
C:Accession: E70390  
R;Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; V. Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98156666; PMID:9537320  
A:Accession: E70390  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-370 <AQF>  
A:Cross-references: GB:AE000720; NID:G2983529; PIDN:AAC07111.1; PID:G2983536; GB:AE00065  
A:Experimental source: strain VF5  
C:Genetics:  
A:Gene: nifs1  
C:Superfamily: nitrogen fixation protein nifs  
C:Keywords: phosphoprotein; pyridoxal phosphate; sulfotransferase  
F:195/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

F:318/Active site: Cys (cysteine persulfide intermediate) #status predicted  
Query Match 6.9%; Score 87.5; DB 2; Length 370;  
Best Local Similarity 25.4%; Pred. No. 9.9;  
Matches 49; Conservative 30; Mismatches 91; Indels 23; Gaps 8;  
QY 4 KGEELFTGVV---PILVELD---GDVNGHKF-SVSGEG-----EGDATYVGLTLKFICT 50  
Db 164 KGVPLLTDAVQAIGKPIELKNISVATFSGHKFHAIKGSGFLYISDEANVEPLIVGGQE 223  
QY 51 TGKLP-----VPMPTLVTTLSYGVQCFSRYPDHMKQ-HDFFKSAMPEGYVQERTIFFKDD 104  
Db 224 NKGSGTENVVGLSLAKALEIVSNFSYQQLKRLDLFENLLLEA-LPDAQIVGKDA 282  
QY 105 GNYKTRAEV---KPEGDTLVNRIELKIGDFKEDGNTILGHKLEYNNSHNVYIMADKQNG 161  
Db 283 ERSPTSISVIMPKFFGAIEIVNKLSEKGIYCSGACLSGEYBPKNHMLKMGFSQEKALRM 342  
QY 162 IKVNFKIRHNIED 174  
Db 343 VRFSFGLLNKEEE 355  
RESULT 10  
AC0582  
leucyl-tRNA synthetase [imported] - Salmonella enterica subsp. enterica serovar Typhi (st  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: this species has also been called Salmonella typhi  
C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: AC0582  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P. Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AC0582  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-860 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD05125.1; PID:G16501899; GSPDB:GN00176  
C:Genetics:  
A:Gene: STY0699  
C:Superfamily: leucine-tRNA ligase

Query Match 6.9%; Score 87.5; DB 2; Length 860;  
Best Local Similarity 23.3%; Pred. No. 30;  
Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;  
QY 50 TTGKLPVPMPTLVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKT 109  
Db 314 TGEIIPV-WAANFVLMYGTGAVMVPGH-DQRD-YEFASKYGLTIKPVILLADGSEPD 370  
QY 110 RAEVFEFGDTLVNRIELKIGDFKEDGNTILGHKLEYNNSHNVYIMADKQNGIKVNFKR 169  
Db 371 SEQALTEKGVLFNSGEFGDLAEAFNAIDKL-----AEKGVGERKVNRLR 418  
QY 170 H-----NTEDSVQLADHYQNTPIGDGPVLLPDNHYL-STQSALSADP 212  
Db 419 DWGVSQRQYWGAPIPMVTLDEDTGTV-----LPTPEDQLPVLIPEDVVMVGITSPKADP 471  
RESULT 11  
S29043  
cellulase (EC 3.2.1.4) - Bacillus sp.  
N:Alternate names: endo-1,4-beta-glucanase  
C:Species: Bacillus sp.  
C>Date: 25-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 01-Dec-2000  
C:Accession: S29043; PC4404  
R;Ozaki, K.; Shikata, S.; Kawai, S.; Ito, S.; Okamoto, K. J. Gen. Microbiol. 136, 1327-1334, 1990  
A:Title: Molecular cloning and nucleotide sequence of a gene for alkaline cellulase from





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:42:24 ; Search time 6.44444 Seconds  
(without alignments)  
1931.085 Million cell updates/sec

Title: US-09-887-784-222I

Perfect score: 1274

Sequence: 1 MVSKGELFTGVVPIVLVD.....VLIGFVTRAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1245	97.7	238	1	GFP_AEQVI
2	91.5	7.2	861	1	SYL_HAEN
3	89.5	7.0	879	1	SYL_XYLFA
4	88.5	6.9	795	1	D152_HAEN
5	88.5	6.9	797	1	D151_HAEN
6	87.5	6.9	793	1	D153_HAEN
7	87.5	6.9	860	1	SYL_SALTI
8	87.5	6.9	860	1	SYL_SALTY
9	87.5	6.9	941	1	GUN_BAGS6
10	87.5	6.9	2222	1	DPOE_YEAST
11	87	6.8	689	1	AC21_HUMAN
12	86.5	6.8	533	1	CP51_CANGA
13	86.5	6.8	879	1	SYL_XYLFT
14	86.5	6.8	1603	1	VIT4_CAEEL
15	86	6.8	357	1	TRMA_CAMJE
16	85.5	6.7	788	1	DPOE_HBHE
17	85.5	6.7	886	1	ITH3_MSAU
18	85	6.7	874	1	SLAP_BACLI
19	84.5	6.6	501	1	AMPA_WIGBR
20	84.5	6.6	613	1	PEPF_MYCPU
21	84.5	6.6	859	1	SYL_SHON
22	84	6.6	353	1	HIS7_BUCAI
23	84	6.6	366	1	SET1_HUMAN
24	84	6.6	461	1	PSBC_CYPAP
25	84	6.6	504	1	YC03_KLEPN
26	83.5	6.6	538	1	GRBE_RAT
27	83.5	6.6	1164	1	BAG_STRAG
28	83	6.5	439	1	SY62_DISOM
29	82.5	6.5	533	1	NIFD_CLOPA
30	82	6.4	682	1	PRC_ECOLI
31	82	6.4	752	1	NECI_RAT
32	82	6.4	774	1	AMY2_SCHPO
33	81.5	6.4	599	1	SYD_HAEDU

RESULT 1  
GFP\_AEQVI  
ID \_GFP\_AEQVI STANDARD; PRT; 238 AA.  
AC P42212; Q17104; Q27903;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Green fluorescent protein.  
GN GFP.  
OS Aequorea victoria (Jellyfish).  
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;  
OC Aequoreidae; Aequorea.  
OX NCBI\_TaxID=6100;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=92175527; PubMed=1347277;  
RA Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,  
RA Cormier M.J.;  
RT "Primary structure of the Aequorea victoria green-fluorescent  
RT protein.";  
RL Gene 111:229-233(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94185810; PubMed=8137953;  
RA Inouye S., Tsuji F.I.;  
RT "Aequorea green fluorescent protein. Expression of the gene and  
RT fluorescence characteristics of the recombinant protein.";  
RN FEBS Lett. 341:277-280(1994).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97299832; PubMed=9154981;  
RA Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;  
RT "Enhanced expression in tobacco of the gene encoding green fluorescent  
RT protein by modification of its codon usage.";  
RL Plant Mol. Biol. 33:989-999(1997).  
RN [4]  
RP CHROMOPHORE.  
RX MEDLINE=93192221; PubMed=8448132;  
RA Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;  
RT "Chemical structure of the hexapeptide chromophore of the Aequorea  
RT green-fluorescent protein.";  
RL Biochemistry 32:1212-1218(1993).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
RX MEDLINE=96355665; PubMed=8703075;  
RA Ormoe M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,  
RA Remington S.J.;  
RT "Crystal structure of the Aequorea victoria green fluorescent  
RT protein.";  
RL Science 273:1392-1395(1996).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
RX MEDLINE=98294543; PubMed=9631087;  
RA Yang F., Moss L.G., Phillips G.N. Jr.;  
RT "The molecular structure of green fluorescent protein.";  
RL Nat. Biotechnol. 14:1246-1251(1996).

## ALIGNMENTS

000116 homo sapien  
Q99nb1 mus musculus  
Q8xbn8 escherichia  
Q8fjy9 escherichia  
P07813 escherichia  
Q58743 methanococc  
Q08684 chlamydomon  
P91679 drosophila  
Q9pkt7 chlamydia m  
Q25443 helicobacte  
P94126 azorhizobiu  
P36924 bacillus ce

34 81.5 6.4 658 1 ADAS\_HUMAN  
35 81 6.4 682 1 AC2L\_MOUSE  
36 80.5 6.3 860 1 SYL\_ECO57  
37 80.5 6.3 860 1 SYL\_ECOL6  
38 80.5 6.3 860 1 SYL\_ECOLI  
39 80 6.3 336 1 YD48\_METJA  
40 80 6.3 461 1 PSBC\_CHLEU  
41 80 6.3 737 1 OPT1\_DROME  
42 79.5 6.2 312 1 TRXB\_CHLMU  
43 79.5 6.2 393 1 TRNB\_HELPY  
44 79.5 6.2 468 1 GLNA\_AZOCA  
45 79.5 6.2 546 1 AMYB\_BACCE



```
Db 181 HYQONTPIGDPVLLPDNHYLSQTSALSKDPNEKRDHIMVLLFVTAAGITHGMDELYK 238
RESULT 2
SVL_HAEIN
ID_SVL_HAEIN STANDARD; PRT; 861 AA.
AC P43827;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--trNA ligase) (LeuRS).
GN LEUS OR H10921.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Karavag A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley K., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd."
RL Science 269:496-512(1995).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32774; AAC22581.1; -.
CC F1R; H64102; H64102.
CC TIGR; H10921; -.
CC HAMAP; MF_00049; -.
CC InterPro; IPR002302; Leu-trNASynt1a.
CC InterPro; IPR002300; tRNA-synt_1a.
CC InterPro; IPR001432; tRNA-synt_1.
CC InterPro; IPR009008; ValRS_fierS_edit.
CC Pfam; PF00133; tRNA-synt_1; 1.
CC PRINTS; PR00985; TRNASYNTHLEU.
CC TIGRFAMS; TIGR00396; leuS_bact; 1.
CC PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT BINDING 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
SQ SEQUENCE 861 AA; 97750 MW; EB93304F6B4C8FE7 CRC64;
Query Match 7.2%; Score 91.5; DB 1; Length 861;
Best Local Similarity 24.1%; Pred. No. 7;
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;
QY 50 TTGKLPVPTLVTLSYGVCFSRYPDHMKQHDFFKSAMPEGYQVERIFFKD-----103
DB 314 TGDKLPI-WVANFVLMHYGTGMVAFPAH-DQRDF-----EFAQKYSIPKIQIAPLA 364
```

```
QY 104 DGNKYTRAEVKFEGDTLVNRIELKGDIFKEDGNILGHKLEYNNYNNHYIMADK-QKNGI 162
DB 365 DBEIDLTKQAFVEHGKLVNSDFDGKNF--DGAFFNG-----IADKLEKLGV 408
QY 163 ---KVNFKIRH-----NIEDSQVLADHYQQNTPIGDPVLLPDNHYL- 202
DB 409 GKRQVNYRLRDWGVSRQRYWGAPIPMLTLENGDVVFA-----PMEDUPFIILPEDVVM 461
QY 203 STQSALSKDPN 213
DB 462 GYKSPINADPN 472
RESULT 3
SVL_XYLFA
ID_SVL_XYLFA STANDARD; PRT; 879 AA.
AC Q9PEG8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--trNA ligase) (LeuRS).
GN LEUS OR XF2176.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=9a5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facinanci A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Fuzlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa."
RL Nature 406:151-159(2000)
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF004031; AAF84975.1; ALT_INIT.
CC HAMAP; MF_00049; -.
DR
```





CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; U13961; AAA85645.1; -;  
 DR EMBL; U60832; AAB61974.1; -;  
 DR EMBL; U60833; AAB61976.1; -;  
 DR PIR; JC4078; JC4078.

DR InterPro; IPR000184; Bac surfAg\_D15.

DR Pfam; PF01103; Bac surface Ag; 1.

KW Antigen; Outer membrane; Signal.

FT SIGNAL 1 19 POTENTIAL.

FT CHAIN 20 797 PROTECTIVE SURFACE ANTIGEN D15.

SQ SEQUENCE 797 AA; 87675 MW; 2F93DES3869AFIB CRC64;

Query Match 6.9%; Score 88.5; DB 1; Length 797;  
 Best Local Similarity 21.9%; Pred. No. 11;  
 Matches 48; Conservative 30; Mismatches 78; Indels 63; Gaps 11;

QY 65 LSYGVQCFRSYDPHMKQHDF-----FKSAMPEGYVQE-----RTI 99

Db 427 IGYGTESGISYQASVKQDNFLGTGAASVSTAGTKNDYGTSVNLGYTEPYTKDGVSLGGNV 486

QY 100 FFKDDGNKYTRAEVKFEGDTLVNRIELKIDPKEDGNI---LGH-----KLEYNYS 148

Db 487 PFENYDNSKSDTSSNRYKRTTYGNSVTL-GFPVNNNSYYVGLGHTYKNISNFALEYN--- 542

QY 149 HNVYIMADKQK-NGIKVNFKIRHNIEDGVSQVLADHYQO-----NTPIGDGPVLL 196

Db 543 RNLXTQSMKFKNGIKTN-----DFDFSGWYNNSLNRGYFPTKGVKASLG-GRVTI 593

QY 197 P--DNHYLSTQSALSADVPNEKRDHVMVLIGFVTAAGITLG 233

Db 594 PGSDNKYKLSADVQGFPLDRDRLHVMVWSAKASAGYANG 632

RESULT 6

ID D153 HAEIN STANDARD; PRT; 793 AA.

AC Q32629;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE Protective surface antigen D15 precursor (80 kDa D15 antigen)

DE (D-15-Ag) (Outer membrane protein D15).

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

OC Pasteurellaceae; Haemophilus.

OX NCBI\_TaxID=727;

RN [1]

RC SEQUENCE FROM N.A.

RA STRAIN=PAK 12085;

RX MEDLINE=97427952; PubMed=9284140;

RA Loomore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,

RA Klein M.H.;

RT "Outer membrane protein D15 is conserved among Haemophilus influenzae

RT species and may represent a universal protective antigen against

RT invasive disease."

RL Infect. Immun 65:3701-3707(1997).

CC -!- SUBCELLULAR LOCATION: Outer membrane.

CC -!- SIMILARITY: Belongs to the surface antigen D15 family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U60834; AAB61977.1; -;

DR InterPro; IPR000184; Bac surfAg\_D15.

DR Pfam; PF01103; Bac surface Ag; 1.

KW Antigen; Outer membrane; Signal.

FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 793 PROTECTIVE SURFACE ANTIGEN D15.  
 SQ SEQUENCE 793 AA; 87511 MW; 51BFDB2036801A14 CRC64;

Query Match 6.9%; Score 87.5; DB 1; Length 793;  
 Best Local Similarity 22.4%; Pred. No. 13;  
 Matches 49; Conservative 28; Mismatches 79; Indels 63; Gaps 11;

QY 65 LSYGVQCFRSYDPHMKQHDF-----FKSAMPEGYVQE-----RTI 99

Db 427 IGYGTESGISYQTSIKQDNFLGTGAASVSTAGTKNDYGTSVNLGYTEPYTKDGVSLGGNI 486

QY 100 FFKDDGNKYTRAEVKFEGDTLVNRIELKIDPKEDGNI---LGH-----KLEYNYS 148

Db 487 PFENYDNSKSDTSSNRYKRTTYGNSVTL-GFPVNNNSYYVGLGHTYKNISNFALEYN--- 542

QY 149 HNVYIMADKQK-NGIKVNFKIRHNIEDGVSQVLADHYQO-----NTPIGDGPVLL 196

Db 543 RNLXTQSMKFKNGIKTN-----DFDFSGWYNNSLNRGYFPTKGVKASLG-GRVTI 593

QY 197 P--DNHYLSTQSALSADVPNEKRDHVMVLIGFVTAAGITLG 233

Db 594 PGSDNKYKLSADVQGFPLDRDRLHVMVWSAKASAGYANG 632

RESULT 7

ID SYL\_SALTI STANDARD; PRT; 860 AA.

AC Q8Z8H5;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).

GN LEUS OR STY0659 OR T2219.

OS Salmonella typhi.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Salmonella.

OX NCBI\_TaxID=601;

RN [1]

RC SEQUENCE FROM N.A.

RA STRAIN=CT18;

RX MEDLINE=21534947; PubMed=11677608;

RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,

RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,

RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,

RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Barrar J.,

RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,

RA Krogh A., Larsen T.S., Leather S., Moule S., O'Goara P., Parry C.,

RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,

RA Whitehead S., Barrell B.G.;

RT "Complete genome sequence of a multiple drug resistant Salmonella

RT enterica serovar Typhi CT18."

RL Nature 413:848-852(2001).

RN [2]

RC SEQUENCE FROM N.A.

RA STRAIN=Ty2 / ATCC 700931;

RX MEDLINE=22531367; PubMed=12644504;

RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,

RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;

RT "Comparative genomics of Salmonella enterica serovar typhi strains Ty2

RT and CT18."

RL J. Bacteriol. 185:2330-2337(2003).

CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +

CC diphosphate + L-leucyl-tRNA(Leu).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.

CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL627267; CAD05125.1; -.
DR EMBL; AE016841; AA069822.1; -.
DR HAMAP; MF 00049; -.
DR InterPro; IPR002302; Leu-TRNAsynt1a.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR009009; ValRS_1IERS_edit.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRFAMs; TIGR00396; leus bact; 1.
DR PROSITE; PS00178; AA TRNA LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
SQ SEQUENCE 860 AA; 96940 MW; 2F95E480BBAB23C4 CRC64;

Query Match 6.9%; Score 87.5; DB 1; Length 860;
Best Local Similarity 23.3%; Pred. No. 15;
Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;

QY 50 TTGKLPVPWPTLVTTLSYGVCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNYKT 109
Db 314 TGEIIPV-WAANFVMEYGTGAVMVPCH-DQRD-YEFASKYGLTIKPVILAADGSEPD 370
QY 110 RAEVKFGDPTLVNRIELKGIDFKEDGNILGHKLEYNVSHVYIMADKQNGIKVNFKIR 169
Db 371 SEQALTEKGVLFNSGFEFDGLAFEAFAFNAIADKL-----AEKGVGERKVNRYLR 418
QY 170 H-----NIEDGSVOLADHYQNTPIGDGPVLLPDNHYL-STQSALS KDP 212
Db 419 DMGVSQRQYWGAPIPWVLTEDGTV-----LPTPDQLPVLPEDVMDGITSPIKADP 471

RESULT 8
SYL_SALTY
ID_SYL_SALTY STANDARD; PRT; 860 AA.
AC Q8ZQ26;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR STM0648.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SCSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2."
RL Nature 413:852-856 (2001).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M27420; AAA22304.1; -.
DR PIR; S29043; S29043.
DR PDB; 1G01; 31-DEC-02.
```

```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE008725; AAL19599.1; -.
DR StyGene; SG????; leus.
DR HAMAP; MF 00049; -.
DR InterPro; IPR002302; Leu-TRNAsynt1a.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR009008; ValRS_1IERS_edit.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRFAMs; TIGR00396; leus bact; 1.
DR PROSITE; PS00178; AA TRNA LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
SQ SEQUENCE 860 AA; 96985 MW; D5003584DFCCAB6 CRC64;

Query Match 6.9%; Score 87.5; DB 1; Length 860;
Best Local Similarity 23.3%; Pred. No. 15;
Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;

QY 50 TTGKLPVPWPTLVTTLSYGVCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNYKT 109
Db 314 TGEIIPV-WAANFVMEYGTGAVMVPCH-DQRD-YEFASKYGLTIKPVILAADGSEPD 370
QY 110 RAEVKFGDPTLVNRIELKGIDFKEDGNILGHKLEYNVSHVYIMADKQNGIKVNFKIR 169
Db 371 SEQALTEKGVLFNSGFEFDGLAFEAFAFNAIADKL-----AEKGVGERKVNRYLR 418
QY 170 H-----NIEDGSVOLADHYQNTPIGDGPVLLPDNHYL-STQSALS KDP 212
Db 419 DMGVSQRQYWGAPIPWVLTEDGTV-----LPTPDQLPVLPEDVMDGITSPIKADP 471

RESULT 9
GUN_BACS6
ID_GUN_BACS6 STANDARD; PRT; 941 AA.
AC P19424;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Alkaline cellulase).
OS Bacillus sp. (strain KSM-635).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1415;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91037937; PubMed=2230718;
RA Ozaki K., Shikata S., Kawai S., Ito S., Okamoto K.;
RT "Molecular cloning and nucleotide sequence of a gene for alkaline cellulase from Bacillus sp. KSM-635."
RL J. Gen. Microbiol. 136:1327-1334 (1990).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -1- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC -1- SIMILARITY: Contains 3-layer homology (SLH) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M27420; AAA22304.1; -.
DR PIR; S29043; S29043.
DR PDB; 1G01; 31-DEC-02.
```

```
DR PDB; 1GOC; 31-DEC-02.
DR InterPro; IPR005086; CBM_17_28.
DR InterPro; IPR008979; Gal_bind like.
DR InterPro; IPR001547; Glyco_hydro_5.
DR InterPro; IPR001119; SLH.
DR Pfam; PF03424; CBM_17_28; 2.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00395; SLH; 3.
DR PROSITE; PS00659; GLYCOSYL HYDROL_F5; 1.
DR PROSITE; PS01072; SLH_DOMAIN; 2.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat;
KW 3D-structure.
FT SIGNAL 1 29
FT CHAIN 30 941 ENDOGLUCANASE.
FT DOMAIN 40 99 SLH 1.
FT DOMAIN 100 151 SLH 2.
FT DOMAIN 152 225 SLH 3.
FT ACT_SITE 373 373 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 485 485 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 941 AA; 104628 MW; BEA2AC3B169BFADA CRC64;

Query Match 6.9%; Score 87.5; DB 1; Length 941;
Best Local Similarity 20.3%; Pred. No. 16;
Matches 49; Conservative 33; Mismatches 62; Indels 97; Gaps 11;

QY 16 LVELDGVNGHKFSVSGEGEDATYGLKTLKFTCTGKLPVPMPTTLVTLTSLVGVQCFSRY 75
Db 240 LVELNG-----QUTLAGE---DGT-----PVQLRGWST---HGLQWFG-- 271

QY 76 PDHMKQHDFFKSMPSGQYQVQERTIFFKDDGNYKTRAEVKEGDTLVNRLKGLDKIDPKDG 135
Db 272 -ETVNEAFVSLNDWGSNMRLAMVIGENYATNPVK---DLVYEGIELA----- 319

QY 136 NILGHKLEYNNSHNYVIMADKQNGIKVNFKRHNIEDGSV---QLADHYQONTPIGD 191
Db 320 -----FHDYVIYVDWH---VHAPGDPADVSGAYDFFEIADHYKDH----- 360

QY 192 GPVLLPDNHYLSQSALSKDPN-----EKRDHMLVI 222
Db 361 -----PNHYIITWELANFSPNNGGLTNDKGEAVKEVPEIVEMLRKGDNMILV 415

QY 223 G 223
Db 416 G 416

RESULT 10
DPOE_YEAST STANDARD; PRT; 2222 AA.
AC P21951;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA polymerase epsilon, catalytic subunit A (EC 2.7.7.7) (DNA
DE polymerase II subunit A).
GN POL2 OR DUN2 OR YNL262W OR N0825.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1214-1221.
RX MEDLINE=90381771; PubMed=2169349;
RA Morrison A., Araki H., Clark A.B., Hanatake R.K., Sugino A.;
RT "A third essential DNA polymerase in S. cerevisiae.";
RL Cell 62:1143-1151(1990).
RN [2]
RP SEQUENCE OF 1-2221 FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=96310631; PubMed=8740425;
RA Sen-Gupta M., Lyck R., Fiebig U., Niedenthal R.K., Hegemann J.H.;
RT "The sequence of a 24,152 bp segment from the left arm of chromosome
RT XIV from Saccharomyces cerevisiae between the BNI1 and the POL2
```



Db 473 FFGIVPVLMDKSGV-----VEGNSVSGALCIS-----QAWPGMARTI--- 510  
QY 69 VQCFSRYPDHMKHQHDFKSPGEGYQERTIFFKDDGNKYKTRA---EVKFEGLTLVNRIE 125  
Db 511 -----YGDHQRVDYFRKYP-GY-----YFTGDGARTGEGYQITGRMDVVI----- 553  
QY 126 LKGIDPKEDGNILGHKL 142  
Db 554 -----NISGRL 560

RESULT 12  
CP51\_CANGA STANDARD; PRT; 533 AA.  
AC P50859; Q02312;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Cytochrome P450 51 (EC 1.14.13.70) (CYPLI) (P450-LIA1) (Sterol 14-alpha demethylase) (lanosterol 14-alpha demethylase) (P450-14DM).  
GN ERG11 OR CYP51.  
OS Candida glabrata (Yeast) (Torulopsis glabrata).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OX NCBI\_TaxID=5478;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=2001-L5;  
RX MEDLINE=96161286; PubMed=8593007;  
RA Geber A., Hitchcock C.A., Swartz J.E., Pullen F.S., Marsden K.E., Kwon-Chung K.J., Bennett J.E.;  
RT "Deletion of the Candida glabrata ERG3 and ERG11 genes: effect on cell viability, cell growth, sterol composition, and antifungal susceptibility";  
RT Antimicrob. Agents Chemother. 39:2708-2717(1995).  
RN [2]  
RP SEQUENCE OF 60-473 FROM N.A.  
RC STRAIN=ATCC 2001;  
RX MEDLINE=95081364; PubMed=7989540;  
RA Burgener-Kairuz P., Zuber J.P., Jaunin P., Buchman T.G., Bille J., Rossier M.;  
RT "Rapid detection and identification of Candida albicans and Torulopsis (Candida) glabrata in clinical specimens by species-specific nested PCR amplification of a cytochrome P-450 lanosterol-alpha-demethylase (lial) gene fragment";  
RL J. Clin. Microbiol. 32:1902-1907(1994).  
CC -!- FUNCTION: Catalyzes C14-demethylation of lanosterol which is critical for ergosterol biosynthesis. It transforms lanosterol into 4,4'-dimethyl cholesta-8,14,24-triene-3-beta-ol (By similarity).  
CC -!- CATALYTIC ACTIVITY: Obtusifolliol + 3 O(2) + 3 NADPH = 4-alpha-methyl-5-alpha-ergosta-8,14,24(28)-trien-3-beta-ol + formate + 3 NADP(+) + 3 H(2)O.  
CC -!- PATHWAY: Ergosterol biosynthesis.  
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L40389; AAB02329.1; -;  
DR EMBL; S75389; AAB32679.1; -;  
DR InterPro; IPR001128; Cytochrome\_P450.  
DR Pfam; PF00067; P450; 1.  
DR PRINTS; PR00385; P450.  
DR PROSITE; PS00086; CYTOCHROME P450; 1.  
KW Electron transport; Oxidoreductase; Monooxygenase; Membrane; Heme;  
KW Sterol biosynthesis; NADP.  
FT METAL 472 473 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).

FT CONFLICT 64 64 I -> M (IN REF. 2).  
FT CONFLICT 473 473 I -> T (IN REF. 2).  
SQ SEQUENCE 533 AA; 61305 MW; A0506C17507E6EF7 CRC64;  
Query Match 6.8%; Score 86.5; DB 1; Length 533;  
Best Local Similarity 21.4%; Pred. No. 10;  
Matches 44; Conservative 32; Mismatches 81; Indels 49; Gaps 8;  
QY 25 GHKFSVS---GSGEGDATYKGLTKFICTGKLPVWPMTLVTLTSLVGVQCFSRYPDH--M 79  
Db 109 GHEFIFNAKLADVSAEAAYSHL-----TTPVFGKGVYDCPNHRLM 149  
QY 80 KQHDFEFSAM-PEGYV-----QERTIFFKDDGNKYKTRAEVKFEGLTLVNRIELKGIDF 131  
Db 150 EQKKFVKGALTKEAFVRYVPLIAEEIYKYFRSNKFNENNSGIVDVVMSQPEM--TIF 207  
QY 132 KEDGNILGHKLEYNVSHNYIMADKQNGIKVKIRINIEDSGVOLADHYQONTPIGD 191  
Db 208 TASRSLLGKEMRDKLDTPFAYLYSLDKGFTINF--VFPNLPLEHYRKRDHAQQAIS--- 263  
QY 192 GPVLLPDNHYLSTQSALS KDPNEKRD 217  
Db 264 -----GYMSLIKERREKND 278

RESULT 13  
SYL\_XYLFT STANDARD; PRT; 879 AA.  
AC Q87C65;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).  
GN LEUS OR PDI230.  
OS Xylella fastidiosa (strain Temecual / ATCC 700964).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
OC Xanthomonadaceae; Xylella.  
OX NCBI\_TaxID=183190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=2421331; PubMed=12533478;  
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B., Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H., Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R., Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorry H., Tsai S.M., Carreir H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J., Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E., Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M., Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V., da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T., Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sassaki F.T., Sena J.A.D., de Souza A.A., Truffi D., Teukumo F., Yanai G.M., Zaros L.G., Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C., Kitajima J.P.;  
RT "Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of Xylella fastidiosa";  
RL J. Bacteriol. 185:1018-1026(2003).  
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP + diphosphate + L-leucyl-tRNA(Leu).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE012557; AAO29080.1; ALT\_INIT.  
DR HAMAP; MF\_00049; -; 1.

```
DR InterPro; IPR002302; Leu-TRNAsynt1a.
DR InterPro; IPR002300; tRNA-synt 1a.
DR InterPro; IPR001412; tRNA-synt 1.
DR InterPro; IPR009008; ValRS_fiers__edit.
DR Pfam; PF00133; tRNA-synt 1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRFAMs; TIGR00396; leuS bact; 1.
DR PROSITE; PS00178; AA TRNA_LICASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 45 55 "HIGH" REGION.
FT SITE 637 641 "KMSKS" REGION.
FT BINDING 640 640 ATP (BY SIMILARITY).
FT SEQUENCE 879 AA; 99823 MW; 4C2E01B8FDC497E CRC64;
Query Match
Best Local Similarity 22.2%; DB 1; Length 879;
Matches 44; Conservative 28; Mismatches 69; Indels 57; Gaps 10;
QY 50 TTGKLPVPTLVTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNY-- 107
DB 321 TNEQLPV-WVANFVLMAYGTGAVMVGHDQDQEF--ANKYGLPIRQVIALKEPKNQDE 377
QY 108 -----KTRAEVKFEGDTLVNRIELKIDFKEDGNILGHKLEYNYNHNVYI 153
DB 378 SIWEPDVRWDYADKTR---EFE--LINSAPFDGLDYQAFVLAERFE----- 421
QY 154 MADQKXNG-IKNFKIRHNIEDGSVOLADHYQONTPI-----GDGPVLLPDN 199
DB 422 ---RQGRGQRRVYRLR----DMGVSQRQYWGCFPIVYICPTGCAVPVPENQLPVILPEN 474
QY 200 -HYLSTQSALS KDPNEKR 216
DB 475 VAFSGTGSPKTPDEWRK 492
RESULT 14
VIT4_CABEL STANDARD; PRT; 1603 AA.
AC F18947; Q9BPP3;
DT 01-NOV-1990 (Rel. 16, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vitellogenin 4 precursor.
VIT-4 OR F59D8.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE OF 1-282 FROM N.A.
RA Blumenthal T., Spieth J., Zucker E.;
RL Submitted (NOV-1986) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE OF 1-71 FROM N.A.
RX MEDLINE=85269643; PubMed=4022780;
RA Spieth J., Denison K., Kirtland S., Cane J., Blumenthal T.;
RT "The C. elegans vitellogenin genes: short sequence repeats in the
RT promoter regions and homology to the vertebrate genes."
RL Nucleic Acids Res. 13:5283-5295(1985).
CC -1- FUNCTION: Precursor of the egg-yolk proteins that are sources of
CC nutrients during embryonic development (Potential).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Synthesized in Caenorhabditis only by 32 cells
CC building the intestine of adult hermaphroditic individuals; they
CC are cotranslationally secreted into the body cavity and
CC subsequently taken up by the gonad.
CC -1- SIMILARITY: Contains 1 VWFD domain.
```

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AC024137; AAK09074.1; -.
DR EMBL; M11498; AAZ8163.1; -.
DR EMBL; X02754; CAA26531.1; -.
DR PIR; A43084; A43084.
DR WormPep; F59D8.2; CE26817.
DR InterPro; IPR001747; Lipid_transprt_N.
DR Pfam; PF01347; Vitellogenin_N; 1.
DR SMART; SM00638; LFD_N; 1.
DR SMART; SM00216; VWD; 1.
KW Storage protein; Multigene family; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 1603 VWFD.
FT DOMAIN 1308 1455
FT CONFLICT 30 30 Y -> V (IN REF. 3).
FT CONFLICT 169 169 L -> V (IN REF. 2).
FT CONFLICT 183 187 EVAYT -> RSRLH (IN REF. 2).
FT CONFLICT 275 275 T -> S (IN REF. 2).
SQ SEQUENCE 1603 AA; 186307 MW; E303170325BC99BB CRC64;
Query Match
Best Local Similarity 6.8%; DB 1; Length 1603;
Matches 52; Conservative 32; Mismatches 69; Indels 69; Gaps 12;
QY 1 MYSGKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVGLTLKFICTTGKLPVPWPT 60
DB 162 MESDKDSLFFNVHEKTMGDCVE---AYTIVQEG-GKTYTKSVNFDKCIIR-----PE 211
QY 61 LVTTLSTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIF---FKDDG----- 105
DB 212 TAYGLRFGSEC-----KECEKGGQVQPIVYVTFKNEKLQSEVNSIYT 257
QY 106 -----NYKTRAEVKFEGDTLVNRIELKIDFKEDGNILGHKLEYNYNHNVIMAD 156
DB 258 LNVNGQEVVVKSETRAKVTVEESKINR-EIK-----KVGPKKEIIVYSMENKLEIQ 308
QY 157 KQKNG-----IKVNFKIRHNIEDGSVOLADHYQONT 188
DB 309 FYQGDKAENVNPPKALIEIQKV-EQLLEIFRQIQEH-EQNT 348
RESULT 15
TRMA_CAMJF STANDARD; PRT; 357 AA.
ID TRMA_CAMJE
AC Q9PP92;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE tRNA (Uracil-5-)-methyltransferase (EC 2.1.1.35) (tRNA(M-5-U54)-
DE methyltransferase) (RUMT).
GN TRMA OR CJ0831C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
```







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:45:30 ; Search time 30.6667 Seconds  
(without alignments)  
2458.984 Million cell updates/sec

Title: US-09-887-784-222I  
Perfect score: 1274  
Sequence: 1 MVSKGELFTGVVPIVLVD.....VLIGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25: \*  
1: sp\_archaea: \*  
2: sp\_bacteria: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phase: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp\_vertebrate: \*  
14: sp\_unclassified: \*  
15: sp\_virus: \*  
16: sp\_bacteriaph: \*  
17: sp\_archaea: \*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1241	97.4	238	2 Q8GHE2	Q8ghe2 azotobacter
2	1238	97.2	238	5 Q93125	Q93125 aequorea vi
3	1236	97.0	238	2 Q8GHE4	Q8ghe4 azomonas ag
4	1235	96.9	238	2 Q8GHE3	Q8ghe3 azotobacter
5	1203	94.4	238	5 Q17105	Q17105 aequorea vi
6	1188	93.2	238	5 Q17106	Q17106 aequorea vi
7	1083	85.0	238	5 Q8WTC6	Q8wtc6 aequorea ma
8	1079	84.7	238	5 Q8WP95	Q8wp95 aequorea ma
9	1075	84.4	238	5 Q8WTC4	Q8wtc4 aequorea ma
10	1073	84.2	238	5 Q8WTD0	Q8wtcd0 aequorea ma
11	1072	84.1	238	5 Q8WTC8	Q8wtc8 aequorea ma
12	1072	84.1	238	5 Q8WTC9	Q8wtc9 aequorea ma
13	1070	84.0	238	5 Q8WTC7	Q8wtc7 aequorea ma
14	1068	83.8	238	5 Q8WTC5	Q8wtc5 aequorea ma
15	252.5	19.8	225	5 Q95UA7	Q95ua7 montastraea
16	252.5	19.8	225	5 Q7Z0W5	Q7z0w5 montastraea

17	247	19.4	225	5	Q963F5	Q963f5 montastraea
18	244.5	19.2	236	5	Q8T6U0	Q8t6u0 dendroneph
19	242.5	19.0	225	5	Q7Z0W9	Q7z0w9 montastraea
20	240	18.8	225	5	Q8I6J8	Q8i6j8 trachyphyl
21	238.5	18.7	266	5	Q9U6Y3	Q9u6y3 clavularia
22	233	18.3	225	5	Q7Z0W4	Q7z0w4 montastraea
23	232	18.2	224	5	Q8MU48	Q8mu48 montastraea
24	232	18.2	225	5	Q8T5F1	Q8t5f1 montastraea
25	214	16.8	227	5	Q7Z0W6	Q7z0w6 montastraea
26	214	16.8	234	5	Q7Z0W7	Q7z0w7 montastraea
27	214	16.8	239	5	Q8NMA1	Q8nma1 agaricia ag
28	212.5	16.7	259	5	Q8NMA2	Q8nma2 agaricia fr
29	209	16.4	234	5	Q8T5F2	Q8t5f2 montastraea
30	209	16.4	234	5	Q8MU47	Q8mu47 montastraea
31	208.5	16.4	229	5	Q9U6Y6	Q9u6y6 anemonia ma
32	206	16.2	227	5	Q962P9	Q962p9 montastraea
33	206	16.2	227	5	Q7Z0W8	Q7z0w8 montastraea
34	205.5	16.1	232	5	Q9GP15	Q9gp15 anemonia su
35	205.5	16.1	238	5	Q9BLV9	Q9bly9 renilla mue
36	204.5	16.1	214	5	Q86LV7	Q86lv7 meandrina m
37	204	16.0	221	5	Q95P04	Q95p04 gonopora t
38	203.5	16.0	214	5	Q86LV8	Q86lv8 meandrina m
39	202	15.9	227	5	Q95VT0	Q95vt0 montastraea
40	201.5	15.8	225	5	Q9U6Y8	Q9u6y8 discosoma s
41	201.5	15.8	232	5	Q9GZ28	Q9gz28 anemonia su
42	198.5	15.6	222	5	Q7Z168	Q7z168 cerianthus
43	198.5	15.6	225	5	Q8T6T9	Q8t6t9 radianthus
44	197.5	15.5	232	5	Q9U6Y7	Q9u6y7 discosoma s
45	197	15.5	235	5	Q8T5F0	Q8t5f0 scolymia a cu

ALIGNMENTS

RESULT 1

Q8GHE2 ID Q8GHE2 PRELIMINARY; PRT; 238 AA.  
AC Q8GHE2;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Green fluorescence protein.  
GN 2289GFP.  
OS Azotobacter vinelandii.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Azotobacter.  
OX NCBI\_TaxID=354;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM2289;  
RA Koranyi P., Berenyi M., Burg K.;  
RT "Occurrence of green fluorescence protein in diazotrophic bacteria  
RT Azomonas and Azotobacter.";  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF324408; AAN86140.1; .  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR009017; GFP like.  
DR InterPro; IPR000786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFLUORESCENT.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
SQ SEQUENCE 238 AA; 26914 MW; F84840F1F9064018 CRC64;

Query Match 97.4%; Score 1241; DB 2; Length 238;  
Best Local Similarity 97.5%; Pred. No. 3.6e-96;  
Matches 232; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 2 VSKGELFTGVVPIVLVDGVNGHKFVSVSGEGDATYKLTKEICTTGKLPVWPPTL 61  
Db 1 MSKGELFTGVVPIVLVDGVNGHKFVSVSGEGDATYKLTKEICTTGKLPVWPPTL 60  
QY 62 VTTLISVGVQCFRYPDPMKHQDFFKSAMPEGVYQERTIFFKDDGNVYKTRAEVKEFGDTLV 121

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Green fluorescence protein.  
GN 375GFP.  
OS Azomonas agilis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Azomonas.  
OX NCBI\_TaxID=116849;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Koranyi P., Berenyi M., Burg K.;  
RT "Occurrence of green fluorescence protein in diazotrophic bacteria  
RT Azomonas and Azotobacter.";  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF324405; AAN86137.1; -;  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR009017; GFP\_like.  
DR InterPro; IPR00786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFP; 1.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
SQ SEQUENCE 238 AA; 26902 MW; 15FE9B9C5B4F6B89 CRC64;  
Query Match 97.0%; Score 1236; DB 2; Length 238;  
Best Local Similarity 97.1%; Pred. No. 9.6e-96;  
Matches 231; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 2 VSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGDGYKLTGKLPVWPPTL 61  
DB 1 MSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGDGYKLTGKLPVWPPTL 60  
QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFGDTLV 121  
DB 61 VTFSGVGVCFRYPDHMKRHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFGDTLV 120  
QY 122 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFIRHNIEDGSVOLAD 181  
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFIRHNIEDGSVOLAD 180  
QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITLGMDELYK 239  
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITLGMDELYK 238  
RESULT 4  
Q8GHE3 PRELIMINARY; PRT; 238 AA.  
ID Q8GHE3  
AC Q8GHE3;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Green fluorescence protein.  
GN 85GFP.  
OS Azotobacter vinelandii.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Azotobacter.  
OX NCBI\_TaxID=354;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Koranyi P., Berenyi M., Burg K.;  
RT "Occurrence of green fluorescence protein in diazotrophic bacteria  
RT Azomonas and Azotobacter.";  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF324406; AAN86138.1; -;  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR009017; GFP\_like.  
DR InterPro; IPR00786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFP; 1.  
DR ProDom; PD013756; Green fl protein; 1.  
SQ SEQUENCE 238 AA; 26887 MW; E0E1616BD2AF6188 CRC64;  
Query Match 96.9%; Score 1235; DB 2; Length 238;

Db 61 VTTFSYGVCFSRYPDHMKRHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFGDTLV 120  
QY 122 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFIRHNIEDGSVOLAD 181  
Db 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFIRHNIEDGSVOLAD 180  
QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITLGMDELYK 239  
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITLGMDELYK 238  
RESULT 2  
Q93125 PRELIMINARY; PRT; 238 AA.  
AC Q93125;  
DT 01-FEB-1997 (TrEMBLrel. 02, Created)  
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Green fluorescent protein mutant 3.  
GN GFP.  
OS Aequorea victoria (Jellyfish).  
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;  
OC Aequoreidae; Aequorea.  
OX NCBI\_TaxID=6100;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=96305137; PubMed=8707053;  
RX Cormack B.P., Valdivia R.H., Falkow S.;  
RT "FACS-optimized mutants of the green fluorescent protein (GFP).";  
RL Gene 173:33-38(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Cormack B.P., Bertram G., Egerton M., Gow N.A.R., Falkow S.,  
RA Brown A.J.P.;  
RT "Yeast Enhanced Green Fluorescent Protein (yEGFP): a reporter of gene  
RT expression in Candida albicans.";  
RL Microbiology 0:0-0(1996).  
DR EMBL; U73901; AAB18957.1; -;  
DR HSSP; P42212; IBFP.  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR009017; GFP\_like.  
DR InterPro; IPR00786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFP; 1.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
SQ SEQUENCE 238 AA; 26840 MW; A28622809A9DEA60 CRC64;  
Query Match 97.2%; Score 1238; DB 5; Length 238;  
Best Local Similarity 97.1%; Pred. No. 6.5e-96;  
Matches 231; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 2 VSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGDGYKLTGKLPVWPPTL 61  
DB 1 MSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGDGYKLTGKLPVWPPTL 60  
QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFGDTLV 121  
DB 61 VTFSGVGVCFRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFGDTLV 120  
QY 122 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFIRHNIEDGSVOLAD 181  
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKVNFIRHNIEDGSVOLAD 180  
QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITLGMDELYK 239  
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITLGMDELYK 238  
RESULT 3  
Q8GHE4 PRELIMINARY; PRT; 238 AA.  
ID Q8GHE4  
AC Q8GHE4;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)

```
Best Local Similarity 97.1%; Pred. No. 1.2e-95;
Matches 231; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKPICTTGKLPVPWPTL 61
Db 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKPICTTGKLPVPWPTL 60
QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGLTLV 121
Db 61 VTTFSYGVCFSRYPDHMKRHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGLTLV 120
QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQNGIKVNFKIRNIEDGSVOLAD 181
Db 121 NRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQNGIKVNFKIRNIEDGSVOLAD 180
QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDMHVLIGFVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDMHVLLEFVTAAGITHGMDLYK 238

RESULT 5
Q17105
ID Q17105 PRELIMINARY; PRT; 238 AA.
AC Q17105;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83960; CAA58789.1; -.
DR PIR; JS0692; JQ1514.
DR HSSP; P42212; 1GFL.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR00786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748E44 CRC64;

Query Match 94.4%; Score 1203; DB 5; Length 238;
Best Local Similarity 93.7%; Pred. No. 5.6e-93;
Matches 223; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKPICTTGKLPVPWPTL 61
Db 1 MSKGEELFTGVVPIILVELDGVNGQKFSVSGEGDATYGLTLNFICTTGKLPVPWPTL 60
QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGLTLV 121
Db 61 VTTFSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGLTLV 120
QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQNGIKVNFKIRNIEDGSVOLAD 181
Db 121 NRIELKGIDFKEDGNILGHKMEYNYNHNVIMGDKPKNGIKVNFKIRNIKDGSVOLAD 180
QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDMHVLIGFVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDMHVLLEFVTAAGITHGMDLYK 238

RESULT 6
Q17106
ID Q17106 PRELIMINARY; PRT; 238 AA.
AC Q17106;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83959; CAA58789.1; -.
DR PIR; JS0692; JQ1514.
DR HSSP; P42212; 1GFL.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR00786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748E44 CRC64;

Query Match 93.2%; Score 1188; DB 5; Length 238;
Best Local Similarity 92.9%; Pred. No. 1e-91;
Matches 221; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTLKPICTTGKLPVPWPTL 61
Db 1 MSKGEELFTGVVPIILVELDGVNGQKFSVSGEGDATYGLTLKPICTTGKLPVPWPTL 60
QY 62 VTTLSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGLTLV 121
Db 61 VTTFSYGVCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGLTLV 120
QY 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVIMADKQNGIKVNFKIRNIEDGSVOLAD 181
Db 121 NRIELKGIDFKEDGNILGHKMEYNYNHNVIMGDKPKNGIKVNFKIRNIKDGSVOLAD 180
QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDMHVLIGFVTAAGITLGMDELYK 239
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDMHVLLEFVTAAGITHGMDLYK 238

RESULT 7
Q8WTC6
ID Q8WTC6 PRELIMINARY; PRT; 238 AA.
AC Q8WTC6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPxm19uv;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.O., Chen M.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435431; AAL33916.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR00786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
```



RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF435427; AAL33912.1; -  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR009017; GFP like.  
DR InterPro; IPR000786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFP.1.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
SQ SEQUENCE 238 AA; 26997 MW; 5F80A192173CB84D CRC64;

Query Match 84.1%; Score 1073; DB 5; Length 238;  
Best Local Similarity 81.1%; Pred. No. 4.5e-82;  
Matches 193; Conservative 21; Mismatches 24; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLELDGVNKHGKFSVSGEGDGYGKLTFLKFICTTGKLPVWPPTL 61

DB 1 MSKGEELFTGIVPVLELDGVNKHGKFSVSGEGDGYGKLTFLKFICTTGKLPVWPPTL 60

QY 62 VTTLGVGQCFSPYDPMKQHDFFKSPAMPEGYQERTIFFKDDGNKYKTRAEVKFEGDTLV 121

DB 61 VTTLGVIQCFARYPEHMKMNDFFKSPAMPEGYQERTIFFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181

DB 121 NRIELKGMDFKEDGNILGHKLEYNHSHVYIMPDKANGLKVNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSAISKDPNEKRDMHVLIGFVTAAGITLGHMDELYK 239

DB 181 HYQTNVPLGDGPVLIPIINHLYSQTALSKDRNETRDMHVFLEFFSACGHTGHMDELYK 238

RESULT 11

Q8WTC8 PRELIMINARY; PRT; 238 AA.

AC Q8WTC8;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Green fluorescent protein.

GN GFP.

OS Aequorea macrodactyla.

OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;

OC Aequoreidae; Aequorea.

OX NCBI\_TaxID=147615;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=GFPxm163;

RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,

RA Li S.J., Xia N.S.;

RT "Colorful mutants of green fluorescent protein from Aequorea

RT macrodactyla."

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF435429; AAL33914.1; -

DR GO; GO:0006091; P:energy pathways; IEA.

DR InterPro; IPR009017; GFP like.

DR InterPro; IPR000786; Green\_fl\_protein.

DR Pfam; PF01353; GFP; 1.

DR PRINTS; PR01229; GFP.1.

DR ProDom; PD013756; Green\_fl\_protein; 1.

SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;

Query Match 84.1%; Score 1072; DB 5; Length 238;  
Best Local Similarity 81.1%; Pred. No. 5.5e-82;  
Matches 193; Conservative 21; Mismatches 24; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLELDGVNKHGKFSVSGEGDGYGKLTFLKFICTTGKLPVWPPTL 61

DB 1 MSKGEELFTGIVPVLELDGVNKHGKFSVSGEGDGYGKLTFLKFICTTGKLPVWPPTL 60

QY 62 VTTLGVGQCFSPYDPMKQHDFFKSPAMPEGYQERTIFFKDDGNKYKTRAEVKFEGDTLV 121

DB 61 VTTLGVIQCFARYPEHMKMNDFFKSPAMPEGYQERTIFFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181

DB 121 NRIELKGMDFKEDGNILGHKLEYNHSHVYIMPDKANGLKVNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSAISKDPNEKRDMHVLIGFVTAAGITLGHMDELYK 239

DB 181 HYQTNVPLGDGPVLIPIINHLYSQTALSKDRNETRDMHVFLEFFSACGHTGHMDELYK 238

RESULT 12

Q8WTC9 PRELIMINARY; PRT; 238 AA.

AC Q8WTC9;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Green fluorescent protein.

GN GFP.

OS Aequorea macrodactyla.

OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;

OC Aequoreidae; Aequorea.

OX NCBI\_TaxID=147615;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=GFPxm162;

RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,

RA Li S.J., Xia N.S.;

RT "Colorful mutants of green fluorescent protein from Aequorea

RT macrodactyla."

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF435428; AAL33913.1; -

DR GO; GO:0006091; P:energy pathways; IEA.

DR InterPro; IPR009017; GFP like.

DR InterPro; IPR000786; Green\_fl\_protein.

DR Pfam; PF01353; GFP; 1.

DR PRINTS; PR01229; GFP.1.

DR ProDom; PD013756; Green\_fl\_protein; 1.

SQ SEQUENCE 238 AA; 27031 MW; 5F80A19C19DC584D CRC64;

Query Match 84.1%; Score 1072; DB 5; Length 238;  
Best Local Similarity 81.1%; Pred. No. 5.5e-82;

Matches 193; Conservative 21; Mismatches 24; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLELDGVNKHGKFSVSGEGDGYGKLTFLKFICTTGKLPVWPPTL 61

DB 1 MSKGEELFTGIVPVLELDGVNKHGKFSVSGEGDGYGKLTFLKFICTTGKLPVWPPTL 60

QY 62 VTTLGVGQCFSPYDPMKQHDFFKSPAMPEGYQERTIFFKDDGNKYKTRAEVKFEGDTLV 121

DB 61 VTTLGVIQCFARYPEHMKMNDFFKSPAMPEGYQERTIFFQDDGKYKTRGEVKFEGDTLV 120

QY 122 NRIELKGIDFKEDGNILGHKLEYNHSHVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181

DB 121 NRIELKGMDFKEDGNILGHKLEYNHSHVYIMPDKANGLKVNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSAISKDPNEKRDMHVLIGFVTAAGITLGHMDELYK 239

DB 181 HYQTNVPLGDGPVLIPIINHLYSQTALSKDRNETRDMHVFLEFFSACGHTGHMDELYK 238

RESULT 13

Q8WTC7 PRELIMINARY; PRT; 238 AA.

AC Q8WTC7;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Green fluorescent protein.

GN GFP.

OS Aequorea macrodactyla.

OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;

OC Aequoreidae; Aequorea.

OX NCBI\_TaxID=147615;



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:41:49 ; Search time 47.1111 Seconds  
(without alignments)  
1433.395 Million cell updates/sec

Title: US-09-887-784-2221

Perfect score: 1274

Sequence: 1 MVSKGEELFTGVVPIVLVELD.....VLIGFVTAAGITLGMDELTK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: Genesecp1980s:\*
- 2: Genesecp1990s:\*
- 3: Genesecp2000s:\*
- 4: Genesecp2001s:\*
- 5: Genesecp2002s:\*
- 6: Genesecp2003as:\*
- 7: Genesecp2003bs:\*
- 8: Genesecp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1274	100.0	239	5	Aae17518 Enhanced
2	1274	100.0	363	6	AbR40352 Human ami
3	1274	100.0	893	4	Aag65781 Amino aci
4	1274	100.0	1132	4	Aag65782 Amino aci
5	1286	99.4	239	5	Aae17517 Enhanced
6	1263	99.1	239	3	Aab22882 Enhanced
7	1263	99.1	239	3	Aay54349 Amino aci
8	1263	99.1	239	3	Aay79584 EGFP sign
9	1263	99.1	239	4	Aab50804 Jellyfish
10	1263	99.1	239	4	Aab59900 A. victor
11	1263	99.1	239	4	Aab31171 Amino aci
12	1263	99.1	239	5	Aag66198 A. victor
13	1263	99.1	239	5	Abg94444 Protease
14	1263	99.1	239	5	Aae14599 Aequorea
15	1263	99.1	239	6	Aae34958 Aequorea
16	1263	99.1	239	6	Aag79829 Green flu
17	1263	99.1	239	6	AbR83616 Green flu
18	1263	99.1	239	6	Ada38074 Aequorea
19	1263	99.1	239	7	Abu63204 Aequorea
20	1263	99.1	239	7	Adc18358 EGFP (enh
21	1263	99.1	239	7	Abw00914 Aequorea
22	1263	99.1	239	7	Ade28570 Enhanced
23	1263	99.1	246	7	Abm79011 Enhanced
24	1263	99.1	248	5	Aag68319 Jellyfish
25	1263	99.1	259	5	AAU99804 Biomembra

26	1263	99.1	265	2	AAW97451	Aaw97451 Wild-type
27	1263	99.1	268	5	AAU99803	Aau99803 Biomembra
28	1263	99.1	270	5	AAU99802	Aau99802 Biomembra
29	1263	99.1	272	5	AAU99800	Aau99800 Biomembra
30	1263	99.1	273	5	AAU99801	Aau99801 Biomembra
31	1263	99.1	280	5	AAU99807	Aau99807 Biomembra
32	1263	99.1	281	3	AAU50142	Aay50142 Green flu
33	1263	99.1	281	3	AAU24252	Egfp-MODC
34	1263	99.1	281	5	AAU10888	Egfp-MODC
35	1263	99.1	286	7	AdE28562	Egfp/hum
36	1263	99.1	289	7	AdE28564	Egfp/hum
37	1263	99.1	290	7	AdE28568	Egfp/hum
38	1263	99.1	290	7	AdE28566	Egfp/hum
39	1263	99.1	294	3	AAU22860	GFP-DSVD-
40	1263	99.1	294	3	AAU79638	Caspase-3
41	1263	99.1	294	5	ABG94422	Recombina
42	1263	99.1	308	2	AAU42181	Egfp/DRM
43	1263	99.1	320	6	ABR83620	HUB1-GFP
44	1263	99.1	323	3	AAU54359	GFP mutan
45	1263	99.1	323	6	ABR83621	RUB1-GFP

ALIGNMENTS

RESULT 1	
AAE17518	
ID	AAE17518 standard; protein; 239 AA.
XX	AAE17518;
AC	AAE17518;
DT	22-APR-2002 (first entry)
XX	Enhanced F64L-E222G jellyfish green fluorescent protein mutant.
DE	Jellyfish; green fluorescent protein; GFP; protein redistribution;
XX	cellular function; genetic reporter; mutant; Stoke's shift; muten.
KW	Aequorea victoria.
OS	Synthetic.
XX	Key
XX	Misc-difference 65
FT	Location/Qualifiers
FT	/note= "wild type phe substituted with Leu; This
FT	corresponds to position 64 in the wild type protein"
FT	Misc-difference 223
FT	/note= "wild type Glu substituted with Gly; This
FT	corresponds to position 222 in the wild type protein"
XX	WO2001983338-A2.
XX	27-DEC-2001.
XX	18-JUN-2001; 2001WO-EP006848.
XX	19-JUN-2000; 2000DK-00000953.
XX	20-JUN-2000; 2000US-0212681P.
XX	10-MAY-2001; 2001BK-00000739.
XX	10-MAY-2001; 2001US-0290170P.
XX	(BIOI-) BIOIMAGE AS.
XX	Bjorn SP, Pagliaro L, Thastrup O;
XX	WPI; 2002-098224/13.
XX	N-PSDB; AAD28163.
XX	Novel fluorescent protein in in vitro assay for measuring protein kinase
XX	activity or dephosphorylation activity, or for measuring protein
XX	redistribution, has a green fluorescent protein with F64L and E222G
XX	mutation.
XX	Claim 9; Page 37; 41pp; English.

XX The invention relates to a fluorescent protein derived from green  
CC fluorescent protein (GFP) or its analogue. The GFP containing mutations  
CC at F64L and E222G has a bigger compared to other GFP's making it very  
CC suitable for high throughput screening due to better resolution. The  
CC fluorescent protein is useful in *in vitro* assays for measuring protein  
CC kinase activity or dephosphorylation activity, or for measuring protein  
CC reduction. The fluorescent protein is useful in studying cellular  
CC functions in living cells; as protein tags in transgenic animals, living  
CC and fixed cells; organelle tags, secretion marker and genetic integrity.  
CC The fluorescent protein is also useful as a cell or organelle integrity  
CC marker, a marker for changes in cell morphology, as transfection marker,  
CC and as a marker to be used in combination with fluorescence activated  
CC cell sorting (FACS). The novel proteins can also be used as reporters to  
CC monitor live or dead biomass of organisms, such as fungi. The fluorescent  
CC protein is also useful as markers in transcriptional and translational  
CC fusions for performing transposon vector mutagenesis and as a reporter  
CC for bacterial detection. Transposons encoding the fluorescent protein are  
CC useful for screening promoters and for tagging plasmids and chromosomes.  
CC The fluorescent protein engineered into the genome of a phage is useful  
CC for designing diagnostic tool. The present sequence is a DNA encoding  
CC enhanced F64L-E222G jellyfish green fluorescent protein (GFP) mutant  
XX  
SQ Sequence 239 AA;  
Query Match 100.0%; Score 1274; DB 5; Length 239;  
Best Local Similarity 100.0%; Pred. No. 4.8e-123;  
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60  
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRLELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180  
DB 121 VNRLELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
RESULT 2  
ABR40352  
ID ABR40352 standard; protein; 363 AA.  
XX  
AC ABR40352;  
XX  
DT 08-JUL-2003 (first entry)  
XX  
DE Human amino acid sequence SEQ ID NO: 6.  
XX  
KW Human; heterologous conjugate; intracellular protein.  
XX  
OS Homo sapiens.  
OS Aequoria victoria.  
XX  
FN WO2003029827-A2.  
XX  
PD 10-APR-2003.  
XX  
PF 01-OCT-2002; 2002WO-DK000651.  
XX  
PR 01-OCT-2001; 2001DK-00001433.  
PR 11-OCT-2001; 2001US-0328868P.  
XX  
PA (BIOL-) BIOLMAGE AS.  
XX  
PI Terry BR, Nielsen SJ;

XX  
DR WPI: 2003-430211/40.  
DR N-PSDB; ACC72604.  
XX  
PT Novel cell for identifying modulators of protein interaction, contains a  
PT first conjugate comprising anchor protein, second conjugate having type B  
PT interactor protein and third conjugate with detectable group.  
XX  
PS Disclosure; Page 112-113; 118pp; English.  
XX  
CC The invention relates to a novel cell, comprising three heterologous  
CC conjugates (HC), a first HC (HC1) comprising an anchor protein that  
CC specifically binds to an internal structure within the cell conjugated to  
CC an interactor protein (ip) of type A, a second HC (HC2) comprising IP of  
CC type B conjugated to a first protein of interest, and a third HC (HC3)  
CC comprising a second protein of interest conjugated to detectable group.  
CC The cell is useful for detecting if a compound disrupts or induces the  
CC interaction between two intracellular proteins. The cell is also useful  
CC for screening compounds that modulate the interaction between two  
CC intracellular proteins. The present sequence is used in the  
CC exemplification of the invention  
XX  
SQ Sequence 363 AA;  
Query Match 100.0%; Score 1274; DB 6; Length 363;  
Best Local Similarity 100.0%; Pred. No. 8.8e-123;  
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGDATYVKLTLPKFICTTGKLPVPWPT 60  
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRLELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180  
DB 121 VNRLELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
RESULT 3  
AAG5781  
ID AAG5781 standard; protein; 893 AA.  
XX  
AC AAG5781;  
XX  
DT 07-JAN-2002 (first entry)  
XX  
DE Amino acid sequence of HSPDE4A1-E222G fusion protein.  
XX  
KW PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;  
KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;  
KW fusion protein.  
XX  
OS Homo sapiens.  
OS Aequorea victoria.  
XX  
FN WO200179526-A2.  
XX  
PD 25-OCT-2001.  
XX  
PF 11-APR-2001; 2001WO-DK000264.  
XX  
PR 17-APR-2000; 2000DK-00000651.  
PR 29-MAY-2000; 2000DK-00000849.  
XX  
PA (BIOI-) BIOIMAGE AS.  
XX



PI Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;  
PI Praestegaard M;  
XX WPI; 2001-611727/70.  
DR N-PSDB; AA166852.  
XX  
PT Determining if a compound is a dislocator of PDE4 for identifying  
PT compounds for treating CNS and inflammatory disease comprises identifying  
PT compounds which remove PDE4 spots.  
XX  
XX  
PS Example 1; Page 156-160; 160pp; English.  
XX  
CC The invention relates to determining, if a compound, is a dislocator of  
CC PDE4. The method comprises testing if the compound removes PDE4 spots,  
CC which may optionally be induced by a Rolipram-like reference compound,  
CC and testing if it inhibits the catalytic activity of the PDE4, where the  
CC compound is a dislocator of PDE4, if it removes PDE spots and if it does  
CC not inhibit the catalytic activity of PDE4. The method is useful for  
CC identifying compounds useful for the treatment of diseases of the central  
CC nervous system such as depression and for the treatment of inflammatory  
CC disease such as joint inflammation, Crohn's disease, inflammatory bowel  
CC disease, respiratory diseases, chronic obstructive pulmonary disease  
CC (COPD), including asthma, chronic bronchitis, pulmonary emphysema,  
CC endotoxin shock, toxic shock syndrome, systemic lupus erythematosus,  
CC psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV  
CC infection. The use of a reagent that can mimic or reverse the effect of  
CC the compound with affinity for the catalytic site on intracellular  
CC distribution of the PDE for the preparation of a medicament. The present  
CC sequence represents the amino acid sequence of a HSPDE4A1-E222G fusion  
CC protein  
XX  
SQ Sequence 893 AA;  
  
Query Match 100.0%; Score 1274; DB 4; Length 893;  
Best Local Similarity 100.0%; Pred. No. 3.3e-122; Indels 0; Gaps 0;  
Matches 239; Conservative 0; Mismatches 0;  
  
QY 1 MYSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60  
DB 655 MYSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 714  
  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 715 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 774  
  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 775 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 834  
  
QY 181 DHYQNTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVGFVTAAGITLGMDELYK 239  
DB 835 DHYQNTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVGFVTAAGITLGMDELYK 893  
  
RESULT 4  
AAG65782  
ID AAG65782 standard; protein; 1132 AA.  
AC AAG65782;  
XX  
XX  
DT 07-JAN-2002 (first entry)  
XX  
DE Amino acid sequence of HSPDE4A4-E222G fusion protein.  
XX  
XX PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;  
KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;  
KW fusion protein.  
XX  
XX Homo sapiens.  
OS Aequorea victoria.  
XX  
XX WO200179526-A2.  
XX

PD 25-OCT-2001.  
XX  
PF 11-APR-2001; 2001WO-DK000264.  
XX  
PR 17-APR-2000; 2000DK-00000651.  
PR 29-MAY-2000; 2000DK-00000849.  
XX  
PA (BIOI-) BIOIMAGE AS.  
XX  
PI Terry BR, Scudder KM, Bjorn SP, Thastrup O, Almholt DC;  
PI Praestegaard M;  
XX  
XX WPI; 2001-611727/70.  
DR N-PSDB; AA166853.  
XX  
PT Determining if a compound is a dislocator of PDE4 for identifying  
PT compounds for treating CNS and inflammatory disease comprises identifying  
PT compounds which remove PDE4 spots.  
XX  
PS Example 1; Page 162-167; 160pp; English.  
XX  
CC The invention relates to determining, if a compound, is a dislocator of  
CC PDE4. The method comprises testing if the compound removes PDE4 spots,  
CC which may optionally be induced by a Rolipram-like reference compound,  
CC and testing if it inhibits the catalytic activity of the PDE4, where the  
CC compound is a dislocator of PDE4, if it removes PDE spots and if it does  
CC not inhibit the catalytic activity of PDE4. The method is useful for  
CC identifying compounds useful for the treatment of diseases of the central  
CC nervous system such as depression and for the treatment of inflammatory  
CC disease such as joint inflammation, Crohn's disease, inflammatory bowel  
CC disease, respiratory diseases, chronic obstructive pulmonary disease  
CC (COPD), including asthma, chronic bronchitis, pulmonary emphysema,  
CC endotoxin shock, toxic shock syndrome, systemic lupus erythematosus,  
CC psoriasis, bone resorption diseases, reperfusion injury, cancer and HIV  
CC infection. The use of a reagent that can mimic or reverse the effect of  
CC the compound with affinity for the catalytic site on intracellular  
CC distribution of the PDE for the preparation of a medicament. The present  
CC sequence represents the amino acid sequence of a HSPDE4A4-E222G fusion  
CC protein  
XX  
SQ Sequence 1132 AA;  
  
Query Match 100.0%; Score 1274; DB 4; Length 1132;  
Best Local Similarity 100.0%; Pred. No. 4.7e-122;  
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MYSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 60  
DB 894 MYSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGKLTLCFICTTGKLPVPWPT 953  
  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 954 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVQERTIFFKDDGNYKTRAEVKFEGDTL 1013  
  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 1014 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 1073  
  
QY 181 DHYQNTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVGFVTAAGITLGMDELYK 239  
DB 1074 DHYQNTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHMLVGFVTAAGITLGMDELYK 1132  
  
RESULT 5  
AAE17517  
ID AAE17517 standard; protein; 239 AA.  
XX  
AC AAE17517;  
XX  
XX  
DT 22-APR-2002 (first entry)  
XX  
XX Enhanced F64L jellyfish green fluorescent protein mutant.  
DE  
XX

KW	Jellyfish; green fluorescent protein; GFP; protein redistribution;	
KW	cellular function; genetic reporter; mutant; Stoke's shift; muten.	
XX		
OS	Aequorea victoria.	
OS	Synthetic.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 65	
FT	/note= "Wild type Phe substituted with Leu; This	
FT	corresponds to position 64 in the wild type protein"	
FT		
PN	WO200198338-A2.	
XX		
PD	27-DEC-2001.	
XX		
PF	18-JUN-2001; 2001WO-EP006848.	
XX		
PR	19-JUN-2000; 2000DK-00000953.	
PR	20-JUN-2000; 2000US-0212681P.	
PR	10-MAY-2001; 2001DK-00000739.	
PR	10-MAY-2001; 2001US-0290170P.	
XX		
PA	(BIOI-) BIOIMAGE AS.	
XX		
PI	Bjorn SP, Pagliaro L, Thastrup O;	
XX		
DR	WPI; 2002-098224/13.	
DR	N-PSDB; AAD28162.	
XX		
PT	Novel fluorescent protein in in vitro assay for measuring protein kinase	
PT	activity or dephosphorylation activity, or for measuring protein	
PT	redistribution, has a green fluorescent protein with F64L and E222G	
PT	mutation.	
XX		
PS	Example 1; Page 35; 41pp; English.	
XX		
CC	The invention relates to a fluorescent protein derived from green	
CC	fluorescent protein (GFP) or its analogue. The GFP containing mutations	
CC	at F64L and E222G has a bigger compared to other GFP's making it very	
CC	suitable for high throughput screening due to better resolution. The	
CC	fluorescent protein is useful in invitro assays for measuring protein	
CC	kinase activity or dephosphorylation activity, or for measuring protein	
CC	redistribution. The fluorescent protein is useful in studying cellular	
CC	functions in living cells; as protein tags in transgenic animals, living	
CC	and fixed cells; organelle tags, secretion marker and genetic reporter.	
CC	The fluorescent protein is also useful as a cell or organelle integrity	
CC	marker, a marker for changes in cell morphology, as transfection marker,	
CC	and as a marker to be used in combination with fluorescence activated	
CC	cell sorting (FACS). The novel proteins can also be used as reporters to	
CC	monitor live or dead biomass of organisms, such as fungi. The fluorescent	
CC	protein is also useful as markers in transcriptional and translational	
CC	fusions for performing transposon vector mutagenesis and as a reporter	
CC	for bacterial detection. Transposons encoding the fluorescent protein are	
CC	useful for screening promoters and for tagging plasmids and chromosomes.	
CC	The fluorescent protein engineered into the genome of a phage is useful	
CC	for designing diagnostic tool. The present sequence is enhanced F64L	
CC	Jellyfish green fluorescent protein (GFP) mutant	
XX		
SQ	Sequence 239 AA;	
	Query Match 99.4%; Score 1266; DB 5; Length 239;	
	Best Local Similarity 99.6%; Pred. No. 3.2e-122;	
	Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 MVSKEELFTGVPIILVELDGVNKHKFSVSGEGDATYKLTLPKICTTGKLPVWPWT 60	
DB	1 MVSKEELFTGVPIILVELDGVNKHKFSVSGEGDATYKLTLPKICTTGKLPVWPWT 60	
QY	61 LVTTLSYGVQCFSPYDPMKQHDFFKSAMPEGYVQERTIFFKDDGNVKTAEVKEFGDTL 120	
DB	61 LVTTLSYGVQCFSPYDPMKQHDFFKSAMPEGYVQERTIFFKDDGNVKTAEVKEFGDTL 120	
QY	121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180	

Db	121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
QY	181 DHYQONTPIGDGPFVLLPDNHYLSTQSALSKDPEKRDHMLVLFVTAAGITLGMDELYK 239
DB	181 DHYQONTPIGDGPFVLLPDNHYLSTQSALSKDPEKRDHMLVLFVTAAGITLGMDELYK 239

RESULT 6  
AAB22882  
ID AAB22882 standard; protein; 239 AA.  
XX  
AC AAB22882;  
XX  
DT 10-JAN-2001 (first entry)  
XX  
DE Enhanced green fluorescent protein (EGFP), SEQ ID NO:46.  
XX  
KW Bioreactor protein; fusion protein; recognition site;  
KW cellular targeting sequence; cellular localisation; fluorescent protein;  
KW protease activity detection; toxin detection; cellular stress detection;  
KW drug discovery; cell based screening.  
XX  
OS Aequorea victoria.  
OS Synthetic.  
XX  
PN WO2000050872-A2.  
XX  
PD 31-AUG-2000.  
XX  
XX  
PF 25-FEB-2000; 2000WO-US004794.  
XX  
PR 26-FEB-1999; 99US-0122152P.  
PR 08-MAR-1999; 99US-0123399P.  
PR 12-JUL-1999; 99US-00352171.  
XX  
PA (CELL-) CELLOMICS INC.  
XX  
PI Giuliano KA, Kapur R;  
XX  
XX WPI; 2000-594086/56.  
DR N-PSDB; AAA93373.  
XX  
XX Automated cell-based characterization of toxin by contacting cells  
PT containing luminescent reporter molecules with test substance and  
PT analyzing optically.  
XX  
PS Example 11; Fig 29A; 336pp; English.  
XX  
CC The invention relates to systems, methods and reagents for cell-based  
CC screening or detection of compounds which affect particular biological  
CC functions. The methods of the invention utilise fluorescent bioreactor  
CC molecules which, when acted on by a compound of interest, cause an  
CC alteration in the cellular distribution of at least the fluorescent  
CC moiety. In one embodiment, the biosensors comprise heat shock proteins  
CC (HSPs) fused to a fluorescent protein (e.g., jellyfish green fluorescent  
CC protein (GFP), or derivatives thereof). Such biosensors are located in  
CC the cytoplasm, but on stress activation translocate to the nucleus. In  
CC another embodiment bioreactor proteins can be used to detect protease  
CC activity. Such protease bioreactor proteins comprise one or more  
CC fluorescent proteins; a recognition signal which is cleaved by the  
CC protease; and at least one cellular localisation signal. The latter two  
CC components may be components of a single protein which is acted upon by  
CC the protease, or may be from heterologous sources. Due to the  
CC localisation signal, the bioreactor protein is localised to a particular  
CC region of the cell. Once acted on by the protease of interest, the  
CC fluorescent protein is cleaved from the localisation sequence, and is  
CC free to migrate to other locations within the cell. The presence of a  
CC second localisation signal attached to the fluorescent protein enables  
CC the fluorescent protein to be directed to a different cellular  
CC compartment after cleavage of the protease recognition sequence. The  
CC change in distribution of the fluorescent protein can be detected using  
CC imaging methods with a high degree of spatial resolution. The methods and

CC biosensors of the invention can be used to investigate a wide range of  
CC cellular activities and to screen compounds which modulate these  
CC activities. Biosensors containing a recognition site for caspase, for  
CC example, may be used for the screening of compounds which modulate  
CC apoptosis, while biosensors containing other protease recognition sites  
CC may be used for the detection of proteolytic toxins (such as anthrax  
CC lethal factor). The method provides improved target validation and  
CC candidate compound optimisation by combining many cell screening formats  
CC with fluorescence-based molecular reagents and computer-based feature  
CC extraction, data analysis and automation, resulting in increased quantity  
CC and speed of data collection and faster evaluation of drug candidates.  
CC Sequences AAB22881-B22885 represent fluorescent proteins which may used  
CC as components of biosensor fusion proteins of the invention  
XX  
SQ Sequence 239 AA;

Query Match 99.1%; Score 1263; DB 3; Length 239;  
Best Local Similarity 99.2%; Pred. No. 6.6e-122;  
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGKLTLPFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGKLTLPFICTTGKLPVPWPT 60  
QY 61 LVTTLGYVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLGYVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239

RESULT 7  
AA54349  
ID AAY54349 standard; protein; 239 AA.  
XX AAY54349;  
XX  
XX  
XX 06-APR-2000 (first entry)  
DE Amino acid sequence of the mutant green fluorescent protein EGFP.  
XX Fluorescent protein; green fluorescent protein; emission intensity;  
KW fluorescence; pH detection; pH sensor; EGFP.  
OS Synthetic.  
OS Aequorea victoria.

XX Key Location/Qualifiers  
FH Misc-difference 65 /note= "wild type Phe substituted with Leu"  
FT Misc-difference 66 /note= "wild type Ser substituted with Thr"  
FT Misc-difference 232 /note= "wild type His substituted with Leu"  
FT  
XX  
XX WO9964592-A2.  
XX  
XX 16-DEC-1999.  
XX  
XX 08-JUN-1999; 99WO-US012850.  
XX  
XX 09-JUN-1998; 98US-00094359.  
XX 13-OCT-1998; 98US-00172063.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
PA (UYOR-) UNIV OREGON STATE.  
XX

PI Tsien RY, Llopis J, Wachter RM;  
XX WPI; 2000-116540/10.  
DR N-PSDB; AA245642.  
XX  
XX New functional engineered green fluorescent proteins, used for measuring  
XX the pH in biological samples and cells.  
XX  
XX Disclosure; Page 9; 89pp; English.  
PS  
CC The present sequence represents a functional engineered fluorescent  
CC protein based on the Aequorea green fluorescent protein (GFP). The  
CC emission intensity changes as pH varies between 5 and 10 of the present  
CC protein are novel. The functional engineered fluorescent proteins show  
CC reversible changes in fluorescence over physiological pH ranges. They can  
CC be used for determining the pH of samples and cells. The polynucleotides  
CC can also be used to produce transgenic animals. The fluorescent protein  
CC pH sensors can be delivered to cells in the form of polynucleotides  
CC encoding the protein sensor fused to a targeting signal. The targeting  
CC signal directs the expression of the protein sensors to restricted cell  
CC locations. This makes it possible to measure the pH of a precisely  
CC defined cellular region or organelle  
XX  
SQ Sequence 239 AA;

Query Match 99.1%; Score 1263; DB 3; Length 239;  
Best Local Similarity 99.2%; Pred. No. 6.6e-122;  
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGKLTLPFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYGKLTLPFICTTGKLPVPWPT 60  
QY 61 LVTTLGYVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLGYVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLEFVTAAGITLGMDELYK 239

RESULT 8  
AA54349  
ID AAY79584 standard; peptide; 239 AA.  
XX AAY79584;  
XX  
XX 29-AUG-2000 (first entry)  
DE EGFP signal domain.  
XX  
XX Protease; biosensor; EGFP; signal peptide; cell screening; assay;  
KW analysis; drug discovery.  
XX  
XX Unidentified.  
OS  
XX WO200026408-A2.  
XX  
XX 11-MAY-2000.  
XX  
XX 29-OCT-1999; 99WO-US025431.  
XX  
XX 30-OCT-1998; 98US-0106308P.  
XX 26-MAY-1999; 99US-0136078P.  
XX  
XX (CELL-) CELLOMICS INC.  
XX  
XX Guiliano KA, Bright G, Olson K, Burroughs-Tencza S;

XX	WPI; 2000-365644/31.	PF	17-MAY-2000; 2000WO-US013684.
DR	N-PSDB; AAA27573.	XX	
DR		PR	21-MAY-1999; 99US-00316919.
XX		PR	21-MAY-1999; 99US-00316920.
XX	Recombinant nucleic acid encoding a protease biosensor useful for	XX	
PT	fluorescence based cell and molecular biochemical assays for drug	PA	(REGC ) UNIV CALIFORNIA.
PT	discovery comprising three operably linked nucleic acid sequences.	XX	
XX		PI	Tsien RY, Baird GA;
XX	Claim 14; Fig 29A; 218pp; English.	XX	
XX		XX	WPI; 2001-032017/04.
XX	The present sequence is that of the EGFP signal domain, which can be	DR	N-PSDB; AAC90488.
CC	included in novel recombinant protease biosensors (PBs) of the invention.	XX	
CC	The PBs (see AAY79638-54) comprise: a first domain (see AAY79579-87)	PT	Novel fluorescent proteins comprising a sensor protein inserted into
CC	comprising at least 1 detectable polypeptide signal such as the present	PT	them, useful for measuring the response of a sensor biological, chemical,
CC	sequence; a second domain (see AAY79588-622) comprising at least 1	PT	electrical or physiological parameter in vivo or in vitro.
CC	protease recognition site; and a third domain (see AAY79623-37)	XX	
CC	comprising at least 1 reactant target sequence. A recombinant nucleic	PS	Disclosure; Page 24; 94pp; English.
CC	acid (see AAA27627-43) encoding the PB, an expression vector, and a	XX	
CC	genetically engineered host cell are also claimed. A claimed method for	CC	The present sequence is a fluorescent protein used in the construction of
CC	identifying compounds that modify protease activity in a cell involves	CC	a fluorescent protein indicator. The indicator comprises a sensor
CC	contacting a host cell that possesses the recombinant PB with a test	CC	polypeptide that is responsive to a chemical, biological, electrical or
CC	compound, and determining the PB distribution in the host cell, where	CC	physiological parameter, and a fluorescence protein functional group. The
CC	changes in the distribution of the PB are correlated with modification of	CC	sensor polypeptide is operatively inserted into the fluorescent moiety.
CC	protease activity by the test compound. Claimed kits for identifying	CC	The fluorescent indicator is useful for detecting the presence of a
CC	compounds that modify protease activity in a host cell include the	CC	response inducing member in a sample. The method involves contacting the
CC	recombinant nucleic acid, or the recombinant PB, or the vector, or the	CC	sample with the indicator and detecting a change in fluorescence, in
CC	host cell. The PB is useful in high content screens to detect in vivo	CC	which a change is indicative of the effect of the parameter on the sensor
CC	activation of enzymatic activity, and to identify specific activity based	CC	polypeptide. The novel fluorescent proteins are advantageous due to their
CC	on cleavage of a known recognition motif	CC	reduced size as compared to the FRET (fluorescence resonance energy
XX		CC	transfer)-based sensors
SQ	Sequence 239 AA;	XX	
		XX	Sequence 239 AA;
			Query Match 99.1%; Score 1263; DB 4; Length 239;
			Best Local Similarity 99.2%; Pred. No. 6.6e-122;
			Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY	1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTILKFICTTGKLPVPWPT 60	QY	1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTILKFICTTGKLPVPWPT 60
Db	1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTILKFICTTGKLPVPWPT 60	Db	1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTILKFICTTGKLPVPWPT 60
QY	61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120	QY	61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db	61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120	Db	61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY	121 VNRIELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180	QY	121 VNRIELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
Db	121 VNRIELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180	Db	121 VNRIELKGIDFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY	181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239	QY	181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239
Db	181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239	Db	181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239
			RESULT 10
			AAB50804
			ID AAB50804 standard; protein; 239 AA.
XX		XX	
AC	AAB50804;	AC	AAB50804;
XX		XX	
DT	14-MAR-2001 (first entry)	DT	30-NOV-2001 (first entry)
XX		XX	
DE	Jellyfish GFP mutant EGFP.	XX	A. victoria green fluorescent protein (GFP) and linker sequence.
XX		XX	Melanin concentrating hormone receptor; MCHR; MCH; chimeric; fusion;
KW	Aequorea victoria; jellyfish; fluorescent protein indicator;	KW	green fluorescent polypeptide; orexinergic; anabolic; food intake; GFP;
KW	green fluorescent protein; GFP; linker moiety; sensor;	KW	green fluorescent protein.
XX	calmodulin-binding domain; mutant; mutein.	XX	
OS	Aequorea victoria.	OS	Synthetic.
XX		OS	Aequorea victoria.
FN	WO200071565-A2.	XX	
XX		XX	WO200168706-A1.
PD	30-NOV-2000.	XX	
XX		XX	

PD 20-SEP-2001.  
 XX 14-MAR-2001; 2001WO-US008071.  
 PF 15-MAR-2000; 2000US-0189698P.  
 XX (MERI ) MERCK & CO INC.  
 PA Marsh DJ;  
 XX WPI; 2001-565791/63.  
 PI N-PSDB; AAH47304.  
 XX Fusion proteins comprising melanin concentrating hormone receptor  
 PT peptides and fluorescent proteins, useful for identifying appetite  
 PT stimulants.  
 XX  
 PS Claim 2; Page 14; 71pp; English.  
 XX The invention provides melanin concentrating hormone (MCH) receptor  
 CC (MCHR) chimeric and fusion proteins. The MCHR chimeric proteins comprise  
 CC MCHR polypeptide regions from different species. The MCHR fusion protein  
 CC comprise MCHR polypeptide region and a fluorescent polypeptide region  
 CC joined directly, or via a linker, to the carboxy side of the MCHR  
 CC polypeptide region. The MCHR fusion proteins can be expressed by standard  
 CC recombinant methodology. MCH action promotes feeding (orexigenic) and up  
 CC regulation of MCH activity stimulates food intake. The present sequence  
 CC represents a A. victoria green fluorescent protein (GFP) and a linker  
 CC sequence  
 XX  
 SQ Sequence 239 AA;  
 Query Match 99.1%; Score 1263; DB 4; Length 239;  
 Best Local Similarity 99.2%; Pred. No. 6.6e-122;  
 Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYKLTAKFICTTGKLPVWPWT 60  
 DB 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYKLTAKFICTTGKLPVWPWT 60  
 QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
 DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
 RESULT 11  
 AAB31171  
 ID AAB31171 standard; protein; 239 AA.  
 AC AAB31171;  
 XX  
 DT 02-APR-2001 (first entry)  
 XX  
 DE Amino acid sequence of a green fluorescent protein (GFP).  
 XX  
 KW Growth rate; death rate; reporter gene; luminescent protein;  
 KW fluorescent product; luciferase; green fluorescent protein; GFP.  
 XX  
 OS Aequorea victoria.  
 XX  
 PN WO200075367-A1.  
 XX  
 PD 14-DEC-2000.  
 XX  
 PF 07-JUN-2000; 2000WO-FI000507.

XX 07-JUN-1999; 99FI-00001296.  
 PR (LILI/) LILIUS E.  
 PA (VIRT/) VIRT M.  
 XX Lilius E, Virta M;  
 XX WPI; 2001-061737/07.  
 DR N-PSDB; AAC86954.  
 XX  
 PT Assessing growth and death rates of a micro-organism in a desired  
 PT environment, by introducing 2 reporter genes encoding luminescent and  
 PT fluorescent products and detecting luminescent fluorescence.  
 XX  
 PS Disclosure; Page 27; 32pp; English.  
 XX  
 CC The specification describes a method for assessing the growth rate and  
 CC death rate of a micro-organism within a predetermined time period in a  
 CC desired environment. The method comprises introducing at least two  
 CC reporter genes encoding luminescent and/or fluorescent products into the  
 CC micro-organisms, incubating the micro-organism within the desired  
 CC environment, and detecting luminescence and/or fluorescence after a  
 CC predetermined time period. Use of two different markers within a micro-  
 CC organism enables the differentiation between growth and death rates. The  
 CC method is used to assess the growth rate and death rate of a micro-  
 CC organism within a predetermined time period in a desired environment. The  
 CC present sequence represents a green fluorescent protein (GFP), and is  
 CC encoded by a plasmid which encodes luminescent and fluorescent proteins,  
 CC and is used in the method of the invention  
 XX  
 SQ Sequence 239 AA;  
 Query Match 99.1%; Score 1263; DB 4; Length 239;  
 Best Local Similarity 99.2%; Pred. No. 6.6e-122;  
 Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYKLTAKFICTTGKLPVWPWT 60  
 DB 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYKLTAKFICTTGKLPVWPWT 60  
 QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
 QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
 DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239  
 RESULT 12  
 AAG66198  
 ID AAG66198 standard; protein; 239 AA.  
 AC AAG66198;  
 XX  
 DT 17-JUN-2002 (first entry)  
 XX  
 DE A. victoria green fluorescent protein (EGFP).  
 XX  
 KW Cyan-green fluorescent protein; fluorescence; recombinant; GFP;  
 KW green fluorescent protein; EGFP.  
 XX  
 OS Aequorea victoria.  
 XX  
 PN JP2002045189-A.  
 XX  
 PD 12-FEB-2002.  
 XX

PF 04-AUG-2000; 2000JP-00237165.  
XX 04-AUG-2000; 2000JP-00237165.  
PR 04-AUG-2000; 2000JP-00237165.  
XX (RIKA ) RIKAGAKU KENKYUSHO.  
PA WPI; 2002-299190/34.  
XX N-PSDE; ABL40628.  
DR A gene encoding cyan-green fluorescent protein.  
XX Example; Page 14; 20pp; Japanese.  
XX The invention relates to a gene encoding proteins having cyan-green  
XX fluorescence characteristic and having a function of showing stable  
XX fluorescence characteristic in acid region. A method for the preparation  
XX of a cyan-green fluorescent protein is provided which involves a  
XX transformant transformed by a recombinant vector comprising the gene,  
XX where the transformant is cultured and the protein is collected from the  
XX culture. The present sequence represents the A. victoria green  
XX fluorescent protein (EGFP)  
XX Sequence 239 AA;  
XX  
XX Query Match 99.1%; Score 1263; DB 5; Length 239;  
XX Best Local Similarity 99.2%; Pred. No. 6.6e-122;  
XX Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYKLTAFICTTGKLPVWPWT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYKLTAFICTTGKLPVWPWT 60  
QY 61 LVTTLSYGVQCFGRYPDPMKQHDFFKSAPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLSYGVQCFGRYPDPMKQHDFFKSAPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFIRHNIEDGSVOLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFIRHNIEDGSVOLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
RESULT 13  
ABG94444  
ID ABG94444 standard; protein; 239 AA.  
XX AC ABG94444;  
XX 27-NOV-2002 (first entry)  
XX Protease biosensor signal sequence #6.  
XX Detection; classification; identification; toxin detection; protease;  
XX ADP-ribosylating toxin; cytotoxic phospholipase; exfoliative toxin;  
XX toxic threat agent.  
XX Synthetic.  
XX US6416959-B1.  
XX 09-JUL-2002.  
XX 25-FEB-2000; 2000US-00513783.  
XX 27-FEB-1997; 97US-00810983.  
XX 27-FEB-1998; 98US-00031271.  
XX 26-FEB-1999; 99US-0122152P.  
XX 08-MAR-1999; 99US-0123399P.  
XX 12-JUL-1999; 99US-00352171.  
XX 31-AUG-1999; 99US-0151797P.  
PR 17-SEP-1999; 99US-00398965.  
PR 29-OCT-1999; 99US-00430656.  
PR 01-DEC-1999; 99US-0168408P.  
XX (GIUL/) GIULIANO K.  
PA (KAPU/) KAPUR R.  
XX Giuliano K, Kapur R;  
PI WPI; 2002-634730/68.  
XX N-PSDB; ABS71491.  
DR Automated cell-based toxin detection, classification, and/or  
XX identification by treating cells involves use of three classes of  
XX luminescent reporter molecules such as detectors, classifiers or  
XX identifiers.  
XX Example 10; Fig 29A; 214pp; English.  
XX The invention describes methods of automated detection, classification  
XX and identification comprising treating cells containing luminescent  
XX reporter molecules (I) in array of locations with a test substance, where  
XX (I) are detectors, classifiers or identifiers, imaging cells in each  
XX location to obtain luminescent signals and converting optical information  
XX into digital data to interpret presence of toxins in the test substance.  
XX The method are useful for detection of toxins chosen from proteases, ADP-  
XX ribosylating toxins, cytotoxic phospholipases, and exfoliative toxins.  
XX Three classes of cell-based luminescent reporter molecules such as  
XX detectors, classifiers and identifiers are described and serve as  
XX reporters of toxic threat agents. The first two levels of  
XX characterisation ensure a rapid readout of toxin class without  
XX sacrificing the ability to detect many new mutant toxins or dissect  
XX several complex mixtures of known toxins. This is the amino acid sequence  
XX of a protease biosensor related signal sequence used in the cell-based  
XX screening system  
XX Sequence 239 AA;  
XX  
XX Query Match 99.1%; Score 1263; DB 5; Length 239;  
XX Best Local Similarity 99.2%; Pred. No. 6.6e-122;  
XX Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYKLTAFICTTGKLPVWPWT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYKLTAFICTTGKLPVWPWT 60  
QY 61 LVTTLSYGVQCFGRYPDPMKQHDFFKSAPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLSYGVQCFGRYPDPMKQHDFFKSAPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFIRHNIEDGSVOLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFIRHNIEDGSVOLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
RESULT 14  
AAE14599  
ID AAE14599 standard; protein; 239 AA.  
XX AC AAE14599;  
XX 31-MAY-2002 (first entry)  
XX Aequorea victoria enhanced green fluorescent protein.  
XX Mutagenesis; enhanced green fluorescent protein; EGFP; mutant; mutein.  
XX Aequorea victoria.  
XX Synthetic.



Db 181 DHYQQNTPIGDPVLLPDNHYLSTQSALSKDPNEKRDMVLLLEFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 15:54:11  
Job time : 48.1111 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:55 ; Search time 12.7778 Seconds  
(without alignments)  
965.630 Million cell updates/sec

Title: US-09-887-784-222L  
Perfect score: 1274  
Sequence: 1 MWSKGEELFTGVVPIVLVD.....VLlGFVTRAGITLGMDELYK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Issued Patents AA:\*
- 1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep.\*
  - 2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep.\*
  - 3: /cgn2\_6/prodata/2/iaa/6A\_COMB.pep.\*
  - 4: /cgn2\_6/prodata/2/iaa/6B\_COMB.pep.\*
  - 5: /cgn2\_6/prodata/2/iaa/PCTUS\_COMB.pep.\*
  - 6: /cgn2\_6/prodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1263	99.1	239	3	US-09-172-063-3
2	1263	99.1	239	4	US-09-513-783A-46
3	1263	99.1	239	4	US-09-316-919-4
4	1263	99.1	239	4	US-09-602-641-3
5	1263	99.1	239	4	US-09-920-922-2
6	1263	99.1	281	3	US-09-062-102-1
7	1263	99.1	281	4	US-09-364-946-1
8	1263	99.1	294	4	US-09-513-783A-2
9	1263	99.1	323	3	US-09-172-063-21
10	1263	99.1	323	4	US-09-602-641-21
11	1263	99.1	364	3	US-09-085-305-6
12	1263	99.1	379	4	US-09-417-197-129
13	1263	99.1	434	4	US-09-800-170-48
14	1263	99.1	442	4	US-09-417-197-127
15	1263	99.1	459	4	US-09-513-783A-170
16	1263	99.1	544	4	US-09-417-197-113
17	1263	99.1	544	4	US-09-417-197-115
18	1263	99.1	604	4	US-09-417-197-59
19	1263	99.1	605	4	US-09-417-197-41
20	1263	99.1	606	4	US-09-417-197-65
21	1263	99.1	607	4	US-09-417-197-47
22	1263	99.1	630	4	US-09-417-197-63
23	1263	99.1	631	4	US-09-417-197-39
24	1263	99.1	633	4	US-09-417-197-45
25	1263	99.1	635	4	US-09-417-197-125
26	1263	99.1	642	2	US-08-818-253-2
27	1263	99.1	642	2	US-08-818-253-6

ALIGNMENTS

RESULT 1  
US-09-172-063-3  
; Sequence 3, Application US/09172063  
; Patent No. 6150176  
; GENERAL INFORMATION:  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Miyawaki, Atsushi  
; APPLICANT: Lloplis, Juan  
; APPLICANT: Wachter, Rebekka M.  
; APPLICANT: Remington, S. James  
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
; FILE REFERENCE: 07257/071001  
; CURRENT APPLICATION NUMBER: US/09/172,063  
; CURRENT FILING DATE: 1998-10-13  
; EARLIER APPLICATION NUMBER: 09/094,359  
; EARLIER FILING DATE: 1998-06-09  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: EGPP  
US-09-172-063-3

Query Match 99.1%; Score 1263; DB 3; Length 239;  
Best Local Similarity 99.2%; Pred. No. 1.4e-127;  
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MWSKGEELFTGVVPIVLVDGVNGHKFSVSGEGDATYVGKLTAKFICTTGKLPVWPWT 60  
Db 1 MWSKGEELFTGVVPIVLVDGVNGHKFSVSGEGDATYVGKLTAKFICTTGKLPVWPWT 60  
QY 61 LVTLLSYGVQCSRYPDHMKQHDFFKSAPEGVQVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
Db 61 LVTLLTYGVQCSRYPDHMKQHDFFKSAPEGVQVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGDIFKEDGNILGHKLEYNNYNNVIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
Db 121 VNRIELKGDIFKEDGNILGHKLEYNNYNNVIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTOSALS KDPNEKRDHVVLLGFVTAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDGPVLLPDNHYLSTOSALS KDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

Sequence 2, Appli  
Sequence 6, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 75, Appli  
Sequence 51, Appli  
Sequence 71, Appli  
Sequence 139, App  
Sequence 176, App  
Sequence 141, App  
Sequence 143, App  
Sequence 77, Appli  
Sequence 178, App  
Sequence 53, Appli  
Sequence 61, Appli  
Sequence 43, Appli  
Sequence 117, App  
Sequence 119, App

```
RESULT 2
US-09-513-783A-46
; Sequence 46, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGFP
US-09-513-783A-46

Query Match          99.1%; Score 1263; DB 4; Length 239;
Best Local Similarity 99.2%; Pred. No. 1.4e-127;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTLEKFTCTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTLEKFTCTTGKLPVPWPT 60
QY 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSPAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSPAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 3
US-09-316-919-4
; Sequence 4, Application US/09316919
; Patent No. 6469154
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: 07257/073001
; CURRENT APPLICATION NUMBER: US/09/316,919
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-316-919-4

Query Match          99.1%; Score 1263; DB 4; Length 239;
Best Local Similarity 99.2%; Pred. No. 1.4e-127;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTLEKFTCTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTLEKFTCTTGKLPVPWPT 60
QY 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSPAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSPAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 4
US-09-602-641-3
; Sequence 3, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wächter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-602-641-3

Query Match          99.1%; Score 1263; DB 4; Length 239;
Best Local Similarity 99.2%; Pred. No. 1.4e-127;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTLEKFTCTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTLEKFTCTTGKLPVPWPT 60
QY 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSPAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSPAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHVMVLLGFVTAAGITLGMDELYK 239

RESULT 5
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. 6673610
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match          99.1%; Score 1263; DB 4; Length 239;
Best Local Similarity 99.2%; Pred. No. 1.4e-127;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLTKFICTTGKLPVPWPT 60
   |||||
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLTKFICTTGKLPVPWPT 60
   |||||

QY 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGETL 120
   |||||
DB 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGETL 120
   |||||

QY 121 VNRLEKIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
   |||||
DB 121 VNRLEKIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
   |||||

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
   |||||
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
   |||||

RESULT 6
US-09-062-102-1
; Sequence 1, Application US/09062102
; Patent No. 6130313
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangqiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100
; CURRENT APPLICATION NUMBER: US/09/062.102
; EARLIER FILING DATE: 1998-04-17
; EARLIER FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6130313
US-09-062-102-1

Query Match          99.1%; Score 1263; DB 3; Length 281;
Best Local Similarity 99.2%; Pred. No. 1.8e-127;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLTKFICTTGKLPVPWPT 60
   |||||
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLTKFICTTGKLPVPWPT 60
   |||||

QY 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGETL 120
   |||||
DB 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGETL 120
   |||||

QY 121 VNRLEKIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
   |||||
DB 121 VNRLEKIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
   |||||

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
   |||||
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
   |||||

RESULT 7
US-09-364-946-1
; Sequence 1, Application US/09364946
; Patent No. 6306600
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangqiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100CIP/D2
; CURRENT APPLICATION NUMBER: US/09/364.946
; EARLIER FILING DATE: 1999-07-30
; EARLIER FILING DATE: 1999-11-13
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6306600
US-09-364-946-1

Query Match          99.1%; Score 1263; DB 4; Length 281;
Best Local Similarity 99.2%; Pred. No. 1.8e-127;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLTKFICTTGKLPVPWPT 60
   |||||
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLTKFICTTGKLPVPWPT 60
   |||||

QY 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGETL 120
   |||||
DB 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEGETL 120
   |||||

QY 121 VNRLEKIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
   |||||
DB 121 VNRLEKIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
   |||||

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
   |||||
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239
   |||||

RESULT 8
US-09-513-783A-2
; Sequence 2, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513.783A
; EARLIER FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GFP-DEVD-Anexin II construct
US-09-513-783A-2

Query Match          99.1%; Score 1263; DB 4; Length 294;
Best Local Similarity 99.2%; Pred. No. 2e-127;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLTKLTKFICTTGKLPVPWPT 60
   |||||

```

Db 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTILKFICTTGKLPVWPWT 60  
QY 61 LVTLSYGVQCSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDYL 120  
Db 61 LVTTLTYGVQCSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDYL 120  
QY 121 VNRIELKGDIFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
Db 121 VNRIELKGDIFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
QY 181 DHYQNTPIGDPGVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
Db 181 DHYQNTPIGDPGVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 9

US-09-172-063-21  
; Sequence 21, Application US/09172063  
; Patent No. 6150176  
; GENERAL INFORMATION:  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Miyawaki, Atsushi  
; APPLICANT: Llopis, Juan  
; APPLICANT: Wachter, Rebekka M.  
; APPLICANT: Remington, S. James  
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
; FILE REFERENCE: 07257/071001  
; CURRENT APPLICATION NUMBER: US/09/172,063  
; CURRENT FILING DATE: 1998-10-13  
; EARLIER APPLICATION NUMBER: 09/094,359  
; EARLIER FILING DATE: 1998-06-09  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 21  
; LENGTH: 323  
; TYPE: FRT  
; ORGANISM: Aequorea victoria  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: GT-EGFP  
US-09-172-063-21

Query Match 99.1%; Score 1263; DB 3; Length 323;  
Best Local Similarity 99.2%; Pred. No. 2.3e-127;  
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTILKFICTTGKLPVWPWT 60  
Db 85 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTILKFICTTGKLPVWPWT 144  
QY 61 LVTLSYGVQCSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDYL 120  
Db 145 LVTTLTYGVQCSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDYL 204  
QY 121 VNRIELKGDIFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
Db 205 VNRIELKGDIFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 264  
QY 181 DHYQNTPIGDPGVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
Db 265 DHYQNTPIGDPGVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 323

RESULT 10

US-09-602-641-21  
; Sequence 21, Application US/09602641  
; Patent No. 6608189  
; GENERAL INFORMATION:  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Miyawaki, Atsushi  
; APPLICANT: Llopis, Juan

; APPLICANT: Wachter, Rebekka M.  
; APPLICANT: Remington, S. James  
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
; FILE REFERENCE: 07257/071001  
; CURRENT APPLICATION NUMBER: US/09/602,641  
; CURRENT FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 09/172,063  
; PRIOR FILING DATE: 1998-10-13  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 21  
; LENGTH: 323  
; TYPE: FRT  
; ORGANISM: Aequorea victoria  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: GT-EGFP  
US-09-602-641-21

Query Match 99.1%; Score 1263; DB 4; Length 323;  
Best Local Similarity 99.2%; Pred. No. 2.3e-127;  
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTILKFICTTGKLPVWPWT 60  
Db 85 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYVKLTILKFICTTGKLPVWPWT 144  
QY 61 LVTLSYGVQCSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDYL 120  
Db 145 LVTTLTYGVQCSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNGYKTRAEVKFEGDYL 204  
QY 121 VNRIELKGDIFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 180  
Db 205 VNRIELKGDIFKEDGNILGHKLEYNVSHNVYIMADKQNGIKVNFKIRHNIEDGVSQVLA 264  
QY 181 DHYQNTPIGDPGVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
Db 265 DHYQNTPIGDPGVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 323

RESULT 11

US-09-085-305-6  
; Sequence 6, Application US/09085305  
; Patent No. 6191269  
; GENERAL INFORMATION:  
; APPLICANT: Pollock, Allan  
; APPLICANT: Lovett, David H.  
; APPLICANT: Turck, Johanna  
; TITLE OF INVENTION: Selective Induction of Apoptosis in  
; TITLE OF INVENTION: Malignant Cancer Cells by Delivery of N-Terminal  
; TITLE OF INVENTION: Interleukin-1-Alpha Pro-Piece Polypeptide  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Bozicevic & Reed, LLP  
; STREET: 285 Hamilton Ave, Suite 200  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/085,305  
; FILING DATE: 29-MAY-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:

```
ATTORNEY/AGENT INFORMATION:
NAME: Francis, Carol L
REGISTRATION NUMBER: 36,513
REFERENCE/DOCKET NUMBER: 6510/102US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3400
TELEFAX: 650-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 364 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-085-305-6

Query Match          99.1%; Score 1263; DB 3; Length 364;
Best Local Similarity 99.2%; Pred. No. 2.7e-127;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 126 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVPWPT 185

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 186 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 245

QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 246 VNRLEKGDIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 305

QY 181 DHYQQNTPIGDGVPLLPDNNHSLSTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 239
DB 306 DHYQQNTPIGDGVPLLPDNNHSLSTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 364

RESULT 12
US-09-417-197-129
Sequence 129, Application US/09417197
Patent No. 6518021
GENERAL INFORMATION:
APPLICANT: Ole Thastrup, et al.
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An
FILE REFERENCE: 3759-0110P
CURRENT FILING DATE: 1999-10-07
NUMBER OF SEQ ID NOS: 143
SOFTWARE: Patent in version 3.0
SEQ ID NO 129
LENGTH: 379
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: actin-binding-domain-EGFP fusion
US-09-417-197-129

Query Match          99.1%; Score 1263; DB 4; Length 379;
Best Local Similarity 99.2%; Pred. No. 2.9e-127;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 141 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVPWPT 200

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 201 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 260

QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 261 VNRLEKGDIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 320

ATTORNEY/AGENT INFORMATION:
NAME: Francis, Carol L
REGISTRATION NUMBER: 36,513
REFERENCE/DOCKET NUMBER: 6510/102US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3400
TELEFAX: 650-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 364 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-085-305-6

Query Match          99.1%; Score 1263; DB 4; Length 434;
Best Local Similarity 99.2%; Pred. No. 3.5e-127;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 196 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVPWPT 255

QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 256 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 315

QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 316 VNRLEKGDIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLA 375

QY 181 DHYQQNTPIGDGVPLLPDNNHSLSTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 239
DB 376 DHYQQNTPIGDGVPLLPDNNHSLSTQSALSKDPNEKRDHMLVLFVTAAGITLGMDELYK 434

RESULT 14
US-09-417-197-127
Sequence 127, Application US/09417197
Patent No. 6518021
GENERAL INFORMATION:
APPLICANT: Ole Thastrup, et al.
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An
FILE REFERENCE: 3759-0110P
CURRENT FILING DATE: 1999-10-07
NUMBER OF SEQ ID NOS: 143
SOFTWARE: Patent in version 3.0
SEQ ID NO 127
LENGTH: 442
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: EGFP-RhoA fusion
US-09-417-197-127

Query Match          99.1%; Score 1263; DB 4; Length 442;
Best Local Similarity 99.2%; Pred. No. 3.6e-127;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
```

Db 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDGYGKLTLLKFICTTGKLPVWP 60  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

RESULT 15  
US-09-513-783A-170  
; Sequence 170, Application US/09513783A  
; Patent No. 6416959  
; GENERAL INFORMATION:  
; APPLICANT: Giuliano, Kenneth A.  
; TITLE OF INVENTION: A System for Cell Based Screening  
; FILE REFERENCE: 97-022-L1  
; CURRENT APPLICATION NUMBER: US/09/513,783A  
; CURRENT FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 170  
; LENGTH: 459  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP27  
US-09-513-783A-170

Query Match 99.1%; Score 1263; DB 4; Length 459;  
Best Local Similarity 99.2%; Pred. No. 3.8e-127;  
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDGYGKLTLLKFICTTGKLPVWP 60  
Db 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDGYGKLTLLKFICTTGKLPVWP 60  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGVSQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVVLLGFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:04:03  
Job time : 13.7778 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:54:20 ; Search time 34.7778 Seconds

(without alignments)  
1940.117 Million cell updates/sec

Title: US-09-887-784-222L

Perfect score: 1274  
Sequence: 1 MVSKEELFTGVVPIVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processor: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Published Applications AA:

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pcp.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pcp.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pcp.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pcp.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pcp.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pcp.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pcp.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pcp.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pcp.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pcp.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pcp.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pcp.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pcp.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pcp.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pcp.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pcp.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pcp.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1274	100.0	239	9 US-09-887-784-4	Sequence 4, Appli
2	1274	100.0	239	12 US-10-296-953-4	Sequence 4, Appli
3	1274	100.0	363	14 US-10-270-223-6	Sequence 6, Appli
4	1274	100.0	893	14 US-10-257-909A-30	Sequence 30, Appli
5	1274	100.0	1132	14 US-10-257-909A-32	Sequence 32, Appli
6	1266	99.4	239	9 US-09-887-784-2	Sequence 2, Appli
7	1266	99.4	239	12 US-10-296-953-2	Sequence 2, Appli
8	1263	99.1	239	9 US-09-920-922-2	Sequence 2, Appli
9	1263	99.1	239	9 US-09-999-745-4	Sequence 4, Appli
10	1263	99.1	239	10 US-09-866-538-4	Sequence 4, Appli
11	1263	99.1	239	10 US-09-797-496B-2	Sequence 2, Appli
12	1263	99.1	239	10 US-09-794-308-4	Sequence 4, Appli
13	1263	99.1	239	10 US-09-865-291-4	Sequence 4, Appli
14	1263	99.1	239	12 US-10-457-982-3	Sequence 3, Appli
15	1263	99.1	239	14 US-10-121-258-13	Sequence 13, Appli

16	1263	99.1	239	14	US-10-221-461-7	Sequence 7, Appli
17	1263	99.1	239	14	US-10-100-957A-46	Sequence 46, Appli
18	1263	99.1	239	14	US-10-177-390-2	Sequence 2, Appli
19	1263	99.1	239	14	US-10-338-411-3	Sequence 3, Appli
20	1263	99.1	239	15	US-10-370-570-4	Sequence 4, Appli
21	1263	99.1	239	15	US-10-389-640-3	Sequence 3, Appli
22	1263	99.1	259	14	US-10-314-861-11	Sequence 11, Appli
23	1263	99.1	281	12	US-09-931-232-1	Sequence 1, Appli
24	1263	99.1	288	14	US-10-314-861-37	Sequence 37, Appli
25	1263	99.1	293	14	US-10-314-861-35	Sequence 35, Appli
26	1263	99.1	294	14	US-10-100-957A-2	Sequence 2, Appli
27	1263	99.1	295	14	US-10-314-861-39	Sequence 39, Appli
28	1263	99.1	299	14	US-10-314-861-33	Sequence 33, Appli
29	1263	99.1	305	14	US-10-314-861-31	Sequence 31, Appli
30	1263	99.1	308	14	US-10-033-717-35	Sequence 35, Appli
31	1263	99.1	311	14	US-10-314-861-29	Sequence 29, Appli
32	1263	99.1	320	14	US-10-338-411-11	Sequence 11, Appli
33	1263	99.1	320	15	US-10-389-640-11	Sequence 11, Appli
34	1263	99.1	323	12	US-10-457-982-21	Sequence 21, Appli
35	1263	99.1	323	14	US-10-338-411-7	Sequence 7, Appli
36	1263	99.1	323	14	US-10-338-411-13	Sequence 13, Appli
37	1263	99.1	323	15	US-10-389-640-7	Sequence 7, Appli
38	1263	99.1	323	15	US-10-389-640-13	Sequence 13, Appli
39	1263	99.1	324	14	US-10-314-861-16	Sequence 16, Appli
40	1263	99.1	345	14	US-10-338-411-5	Sequence 5, Appli
41	1263	99.1	345	15	US-10-389-640-5	Sequence 5, Appli
42	1263	99.1	346	14	US-10-338-411-9	Sequence 9, Appli
43	1263	99.1	346	15	US-10-389-640-9	Sequence 9, Appli
44	1263	99.1	359	14	US-10-033-717-33	Sequence 33, Appli
45	1263	99.1	359	14	US-10-033-717-34	Sequence 34, Appli

#### ALIGNMENTS

#### RESULT 1

US-09-887-784-4  
; Sequence 4, Application US/09887784  
; Patent No. US20020177189A1  
; GENERAL INFORMATION:  
; APPLICANT: BJORN, Sara et al  
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS  
; FILE REFERENCE: 3759-0115P  
; CURRENT APPLICATION NUMBER: US/09/887,784  
; CURRENT FILING DATE: 2001-06-19  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 239  
; TYPE: PPT  
; ORGANISM: Aequoria Victoria  
US-09-887-784-4

Query Match 100.0%; Score 1274; DB 9; Length 239;  
Best Local Similarity 100.0%; Pred. No. 2e-124;  
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MVSKEELFTGVVPIVELDGDVNGHKFSVSGEGDATYGKLTCLKFCTTCKLPVPWPT	60
Db	1	MVSKEELFTGVVPIVELDGDVNGHKFSVSGEGDATYGKLTCLKFCTTCKLPVPWPT	60
Qy	61	LVTTLSYGVQCFSRYPDHMKQDFFKSAHPGYQERTIFFKDDGNYKTRAEVKPEGDTL	120
Db	61	LVTTLSYGVQCFSRYPDHMKQDFFKSAHPGYQERTIFFKDDGNYKTRAEVKPEGDTL	120
Qy	121	VNRIELKIDFKEDGNILGHKLEYNHSHNYVIMADKQNGIKVNFKIRHNIEDGSVQLA	180
Db	121	VNRIELKIDFKEDGNILGHKLEYNHSHNYVIMADKQNGIKVNFKIRHNIEDGSVQLA	180
Qy	181	DHYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDMHVLGFTVTAAGITLGMDELYK	239
Db	181	DHYQONTPIGDPVLLPDNHYLSTQSALSQDPNEKRDMHVLGFTVTAAGITLGMDELYK	239

***This Page Blank (uspto)***



RESULT 2

US-10-296-953-4  
; Sequence 4, Application US/10296953  
; Publication No. US20040072995A1  
; GENERAL INFORMATION:  
; APPLICANT: BJORN, SARA P.  
; APPLICANT: PAGLIARO, LEN  
; APPLICANT: THASTRUP, OLE  
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS  
; FILE REFERENCE: PL0095  
; CURRENT APPLICATION NUMBER: US/10/296,953  
; CURRENT FILING DATE: 2002-11-26  
; PRIOR APPLICATION NUMBER: PA 2000 00953  
; PRIOR FILING DATE: 2000-06-19  
; PRIOR APPLICATION NUMBER: 60/212,681  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: 60/290,170  
; PRIOR FILING DATE: 2001-05-10  
; PRIOR APPLICATION NUMBER: PA 2001 00739  
; PRIOR FILING DATE: 2001-05-10  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
IS-10-296-953-4

Query Match 100.0%; Score 1274; DB 12; Length 239;  
Best Local Similarity 100.0%; Pred. No. 2e-124;  
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKPEGDTL 120  
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKPEGDTL 120  
QY 121 VNRLEKIGIDFKEDGNILGHKLEYNNSHNVIIMADKQNGIKVNFKIRHNIEDGSGVOLA 180  
DB 121 VNRLEKIGIDFKEDGNILGHKLEYNNSHNVIIMADKQNGIKVNFKIRHNIEDGSGVOLA 180  
QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239

RESULT 3

S-10-270-223-6  
; Sequence 6, Application US/10270223  
; Publication No. US20030143634X1  
; GENERAL INFORMATION:  
; APPLICANT: BioImage A/S  
; TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR COMPONENTS  
; TITLE OF INVENTION: INTERACTING LIVING CELLS, AND TO EXTRACT QUANTITATIVE INFORMATION RE  
; TITLE OF INVENTION: INTERACTIONS BY FLUORESCENCE REDISTRIBUTION.  
; FILE REFERENCE: 3759-0125P  
; CURRENT APPLICATION NUMBER: US/10/270,223  
; CURRENT FILING DATE: 2002-10-11  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 6  
; LENGTH: 363  
; TYPE: PRT  
; ORGANISM: Aequorea Victoria and Human  
S-10-270-223-6

Query Match 100.0%; Score 1274; DB 14; Length 363;  
Best Local Similarity 100.0%; Pred. No. 3.7e-124;  
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKPEGDTL 120  
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKPEGDTL 120  
QY 121 VNRLEKIGIDFKEDGNILGHKLEYNNSHNVIIMADKQNGIKVNFKIRHNIEDGSGVOLA 180  
DB 121 VNRLEKIGIDFKEDGNILGHKLEYNNSHNVIIMADKQNGIKVNFKIRHNIEDGSGVOLA 180  
QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239

RESULT 4

US-10-257-909A-30  
; Sequence 30, Application US/10257909A  
; Publication No. US20030187056A1  
; GENERAL INFORMATION:  
; APPLICANT: Bernard R. TERRY et al.  
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular  
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes  
; FILE REFERENCE: 3759-0125P  
; CURRENT APPLICATION NUMBER: US/10/257,909A  
; CURRENT FILING DATE: 2002-10-17  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 30  
; LENGTH: 893  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Fusion between Aequorea victoria and human  
US-10-257-909A-30

Query Match 100.0%; Score 1274; DB 14; Length 893;  
Best Local Similarity 100.0%; Pred. No. 1.3e-123;  
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60  
DB 655 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 714  
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKPEGDTL 120  
DB 715 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKPEGDTL 774  
QY 121 VNRLEKIGIDFKEDGNILGHKLEYNNSHNVIIMADKQNGIKVNFKIRHNIEDGSGVOLA 180  
DB 775 VNRLEKIGIDFKEDGNILGHKLEYNNSHNVIIMADKQNGIKVNFKIRHNIEDGSGVOLA 834  
QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 239  
DB 835 DHYQONTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHMLLGFVTAAGITLGMDELYK 893

RESULT 5

US-10-257-909A-32  
; Sequence 32, Application US/10257909A  
; Publication No. US20030187056A1  
; GENERAL INFORMATION:  
; APPLICANT: Bernard R. TERRY et al.  
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular  
; TITLE OF INVENTION: distribution of phosphodiesterase (PDE) enzymes  
; FILE REFERENCE: 3759-0125P  
; CURRENT APPLICATION NUMBER: US/10/257,909A  
; CURRENT FILING DATE: 2002-10-17  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 3.0  
US-10-257-909A-32

**This Page Blank (uspto)**

```
; SEQ ID NO 32
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-32

Query Match          100.0%; Score 1274; DB 14; Length 1132;
Best Local Similarity 100.0%; Pred. No. 1.9e-123;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MWSKGEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
DB 894 MWSKGEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 953

QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 954 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 1013

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDSGVQLA 180
DB 1014 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDSGVQLA 1073

QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 1074 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 1132

RESULT 6
US-09-887-784-2
; Sequence 2, Application US/09887784
; Patent No. US20020177189A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-887-784-2

Query Match          99.4%; Score 1266; DB 9; Length 239;
Best Local Similarity 99.6%; Pred. No. 1.4e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MWSKGEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MWSKGEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60

QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDSGVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDSGVQLA 180

QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 7
US-10-296-953-2
; Sequence 2, Application US/10296953
; Publication No. US20040072995A1
; GENERAL INFORMATION:
```

```
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match          99.4%; Score 1266; DB 12; Length 239;
Best Local Similarity 99.6%; Pred. No. 1.4e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MWSKGEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MWSKGEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60

QY 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDSGVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDSGVQLA 180

QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHNVLLGFVTAAGITLGMDELYK 239

RESULT 8
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. US20020083488A1
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match          99.1%; Score 1263; DB 9; Length 239;
Best Local Similarity 99.2%; Pred. No. 2.8e-123;
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MWSKGEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
DB 1 MWSKGEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTKLFICTTGKLPVPWPT 60
```

QY 61 LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 61 LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 9

US-09-999-745-4  
; Sequence 4, Application US/09999745  
; Patent No. US20020157120A1  
; GENERAL INFORMATION:  
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Baird, Geoffrey  
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS  
; FILE REFERENCE: REGEN1470-1  
; CURRENT APPLICATION NUMBER: US/09/999,745  
; CURRENT FILING DATE: 2001-10-23  
; PRIOR APPLICATION NUMBER: 09/316,920  
; PRIOR FILING DATE: 1999-05-21  
; NUMBER OF SEQ ID NOS: 67  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO: 4  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
US-09-999-745-4

Query Match 99.1%; Score 1263; DB 9; Length 239;  
Best Local Similarity 99.2%; Pred. No. 2.8e-123;  
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
QY 61 LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 61 LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 10

US-09-866-538-4  
; Sequence 4, Application US/09866538  
; Publication No. US20030032088A1  
; GENERAL INFORMATION:  
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA  
; APPLICANT: Tsien, Roger  
; APPLICANT: Campbell, Robert  
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS  
; FILE REFERENCE: REGEN1530-2  
; CURRENT APPLICATION NUMBER: US/09/866,538  
; CURRENT FILING DATE: 2001-05-24  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO: 4  
; LENGTH: 239  
; TYPE: PRT

; ORGANISM: Aequorea victoria  
US-09-866-538-4

Query Match 99.1%; Score 1263; DB 10; Length 239;  
Best Local Similarity 99.2%; Pred. No. 2.8e-123;  
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
QY 61 LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 61 LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 11

US-09-797-496B-2  
; Sequence 2, Application US/09797496B  
; Publication No. US20030049597A1  
; GENERAL INFORMATION:  
; APPLICANT: Simon, Sanford M.  
; APPLICANT: Chen, Yu  
; TITLE OF INVENTION: Chimeric Fluorescent Enzymes and Uses Thereof  
; FILE REFERENCE: 600-1-267  
; CURRENT APPLICATION NUMBER: US/09/797,496B  
; CURRENT FILING DATE: 2002-05-24  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO: 2  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Aequorea victoria green fluorescent protein modified as described  
; OTHER INFORMATION: in specification  
US-09-797-496B-2

Query Match 99.1%; Score 1263; DB 10; Length 239;  
Best Local Similarity 99.2%; Pred. No. 2.8e-123;  
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
Db 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
QY 61 LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 61 LVTTLTYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 12

US-09-794-308-4  
; Sequence 4, Application US/09794308  
; Publication No. US20030170911A1  
; GENERAL INFORMATION:  
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA

APPLICANT: TSUEN, Roger  
APPLICANT: ZACHARIAS, David  
APPLICANT: BAIRD, Geoffrey  
TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS  
FILE REFERENCE: REGEN1530  
CURRENT APPLICATION NUMBER: US/09/794,308  
CURRENT FILING DATE: 2001-02-26  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
US-09-794-308-4

Query Match 99.1%; Score 1263; DB 10; Length 239;  
Best Local Similarity 99.2%; Pred. No. 2.8e-123;  
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTLKFKICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTLKFKICTTGKLPVPWPT 60  
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120  
DB 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120  
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239

RESULT 13  
US-09-865-291-4  
Sequence 4, Application US/09865291  
Publication No. US20030186229A1  
GENERAL INFORMATION:  
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA  
APPLICANT: TSUEN, Roger  
APPLICANT: TING, Alice  
APPLICANT: ZHANG, Jin  
TITLE OF INVENTION: EMISSION RATIONETRIC INDICATORS OF PHOSPHORYLATION  
FILE REFERENCE: REGEN1550  
CURRENT APPLICATION NUMBER: US/09/865,291  
CURRENT FILING DATE: 2001-05-24  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
US-09-865-291-4

Query Match 99.1%; Score 1263; DB 10; Length 239;  
Best Local Similarity 99.2%; Pred. No. 2.8e-123;  
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTLKFKICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTLKFKICTTGKLPVPWPT 60  
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120  
DB 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120  
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
RESULT 14  
US-10-457-982-3  
Sequence 3, Application US/10457982  
Publication No. US20030212365A1  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger Y.  
APPLICANT: Miyawaki, Acsushi  
APPLICANT: Ilopis, Juan  
APPLICANT: Wachter, Rebekka M.  
APPLICANT: Remington, S. James  
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
FILE REFERENCE: 07257/071001  
CURRENT APPLICATION NUMBER: US/10/457,982  
CURRENT FILING DATE: 2003-06-09  
PRIOR APPLICATION NUMBER: US/09/602,641  
PRIOR FILING DATE: 2000-06-22  
PRIOR APPLICATION NUMBER: 09/172,063  
PRIOR FILING DATE: 1998-10-13  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (0)...(0)  
OTHER INFORMATION: EGFP  
US-10-457-982-3

Query Match 99.1%; Score 1263; DB 12; Length 239;  
Best Local Similarity 99.2%; Pred. No. 2.8e-123;  
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTLKFKICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTLKFKICTTGKLPVPWPT 60  
QY 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120  
DB 61 LVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120  
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239  
RESULT 15  
US-10-121-258-13  
Sequence 13, Application US/10121258  
Publication No. US20030059835A1  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger  
APPLICANT: Campbell, Robert  
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT  
TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME  
FILE REFERENCE: UC083.1CP2CP1  
CURRENT APPLICATION NUMBER: US/10/121,258  
CURRENT FILING DATE: 2002-04-10  
PRIOR APPLICATION NUMBER: 09/794,308  
PRIOR FILING DATE: 2001-02-26  
PRIOR APPLICATION NUMBER: 09/866,538  
PRIOR FILING DATE: 2001-05-24

; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Enhanced Green Fluorescent Protein (EGFP)  
US-10-121-258-13

Query Match 99.1%; Score 1263; DB 14; Length 239;  
Best Local Similarity 99.2%; Pred. No. 2.8e-123;  
Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 MVSKEELFTGVVPIILVELDGDVNGHKFVSVEGEGDATYKLTKEICTTGKLPVPWPT 60  
Db 1 MVSKEELFTGVVPIILVELDGDVNGHKFVSVEGEGDATYKLTKEICTTGKLPVPWPT 60  
Qy 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Qy 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKVNFIRHNIEDGSVQLA 180  
Db 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKVNFIRHNIEDGSVQLA 180  
Qy 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLGFVTAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLGFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 16:09:28  
Job time : 35.7778 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:00 ; Search time 10.3333 Seconds  
(without alignments)  
2224.817 Million cell updates/sec

Title: US-09-887-784-222L  
Perfect score: 1274  
Sequence: 1 MVSKGEELFTGVVPILVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1241	97.4	238	1 JQ1514	green-fluorescent
2	1105	8.2	785	2 H72228	hypothetical prote
3	91.5	7.2	861	2 H64102	leucine-tRNA ligase
4	90	7.1	632	2 T06586	DNA-binding protei
5	89.5	7.0	887	2 E82590	leucyl-tRNA synthet
6	88.5	6.9	655	2 D83917	DNA topoisomerase
7	87.5	6.9	370	2 E70390	iron-sulfur cofact
8	87.5	6.9	860	2 AC0582	leucyl-tRNA synthet
9	87.5	6.9	2252	1 A36028	DNA-directed DNA p
10	87.5	6.9	2573	2 D71614	hypothetical prote
11	87	6.8	578	1 I40794	dihydroliopamide d
12	87	6.8	874	2 JC4930	S-layer protein pr
13	86.5	6.8	797	2 JC4078	protective surface
14	86.5	6.8	808	2 F64102	protective surface
15	86	6.8	357	2 G81355	tRNA (uracil-5)-m
16	85.5	6.7	788	1 JDUVLH	DNA-directed DNA p
17	85.5	6.7	889	2 JCV576	inter-alpha-trypsi
18	85.5	6.7	941	2 S29043	cellulase (EC 3.2.
19	85	6.7	281	2 AD2052	hypothetical prote
20	85	6.7	439	2 JH0414	synaptogamin o-p65
21	84.5	6.6	425	2 C97354	hypothetical prote
22	84.5	6.6	613	2 A99552	oligodeopeptidase
23	84	6.6	353	2 E84941	imidazoleglycerol-
24	83.5	6.6	836	1 JDUVLD	DNA-directed DNA p
25	83.5	6.6	1134	1 A60234	IGA FC receptor pr
26	83.5	6.6	1164	1 FCSOAG	IGA FC receptor pr
27	83	6.5	461	2 T06936	photosystem II chl
28	83	6.5	471	2 T27856	hypothetical prote
29	83	6.5	774	2 T39539	alpha-amylase homo

ALIGNMENTS

RESULT 1

JQ1514  
green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)  
C:Species: Aequorea victoria  
C:Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 23-Mar-2001  
C:Accession: JQ1514; PQ0335; S48693; S51330; S51331  
R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.  
Gene 111, 229-233, 1992  
A>Title: Primary structure of the Aequorea victoria green-fluorescent protein.  
A:Reference number: JQ1514; MUID:92175527; PMID:1347277  
A:Accession: JQ1514  
A:Molecule type: DNA  
A:Residues: 1-107, 'S', 109-238 <PRA1>  
A:Cross-references: GB:M62654; NID:g155662; PIDN:AAA27722.1; PID:g155663  
A:Accession: JQ1514  
A:Molecule type: mRNA  
A:Residues: 1-99, 'P', 101-140, 'L', 142-218, 'V', 220-238 <PRA2>  
A:Cross-references: GB:M62653; NID:g155660; PIDN:AAA27721.1; PID:g155661  
A:Accession: PQ0335  
A:Molecule type: protein  
A:Residues: 46-64; 74-122; 132-151; 154-183; 185-200 <PRA3>  
R:Inouye, S.; Tsuji, F.I.  
FEBS Lett. 351, 211-214, 1994  
A>Title: Evidence for redox forms of the Aequorea green fluorescent protein.  
A:Reference number: S48693; MUID:94364470; PMID:8082767  
A:Accession: S48693  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-24, 'Q', 26-156, 'P', 158-171, 'K', 173-238 <INO>  
A:Cross-references: GB:L29145; NID:9606383; PIDN:AAA58246.1; PID:g606384  
R:Watkins, J.N.; Campbell, A.K.  
submitted to the EMBL Data Library, January 1995  
A:Reference number: S51330  
A:Accession: S51330  
A:Molecule type: mRNA  
A:Residues: 1-13, 'V', 15-24, 'O', 26-44, 'N', 46-153, 'G', 155-156, 'P', 158-171, 'K', 173-227, 'R', 175-208, 'Q', 209-238, 'P', 240-254, 'S', 256-264, 'L', 266-274, 'G', 276-284, 'D', 286-294, 'E', 296-304, 'K', 306-314, 'R', 316-324, 'G', 326-334, 'L', 336-344, 'D', 346-354, 'E', 356-364, 'K', 366-374, 'R', 376-384, 'L', 386-394, 'D', 396-404, 'E', 406-414, 'K', 416-424, 'G', 426-434, 'L', 436-444, 'D', 446-454, 'E', 456-464, 'K', 466-474, 'R', 476-484, 'L', 486-494, 'D', 496-504, 'E', 506-514, 'K', 516-524, 'G', 526-534, 'L', 536-544, 'D', 546-554, 'E', 556-564, 'K', 566-574, 'R', 576-584, 'L', 586-594, 'D', 596-604, 'E', 606-614, 'K', 616-624, 'G', 626-634, 'L', 636-644, 'D', 646-654, 'E', 656-664, 'K', 666-674, 'R', 676-684, 'L', 686-694, 'D', 696-704, 'E', 706-714, 'K', 716-724, 'G', 726-734, 'L', 736-744, 'D', 746-754, 'E', 756-764, 'K', 766-774, 'R', 776-784, 'L', 786-794, 'D', 796-804, 'E', 806-814, 'K', 816-824, 'G', 826-834, 'L', 836-844, 'D', 846-854, 'E', 856-864, 'K', 866-874, 'R', 876-884, 'L', 886-894, 'D', 896-904, 'E', 906-914, 'K', 916-924, 'G', 926-934, 'L', 936-944, 'D', 946-954, 'E', 956-964, 'K', 966-974, 'R', 976-984, 'L', 986-994, 'D', 996-1004, 'E', 1006-1014, 'K', 1016-1024, 'G', 1026-1034, 'L', 1036-1044, 'D', 1046-1054, 'E', 1056-1064, 'K', 1066-1074, 'R', 1076-1084, 'L', 1086-1094, 'D', 1096-1104, 'E', 1106-1114, 'K', 1116-1124, 'G', 1126-1134, 'L', 1136-1144, 'D', 1146-1154, 'E', 1156-1164, 'K', 1166-1174, 'R', 1176-1184, 'L', 1186-1194, 'D', 1196-1204, 'E', 1206-1214, 'K', 1216-1224, 'G', 1226-1234, 'L', 1236-1244, 'D', 1246-1254, 'E', 1256-1264, 'K', 1266-1274, 'R', 1276-1284, 'L', 1286-1294, 'D', 1296-1304, 'E', 1306-1314, 'K', 1316-1324, 'G', 1326-1334, 'L', 1336-1344, 'D', 1346-1354, 'E', 1356-1364, 'K', 1366-1374, 'R', 1376-1384, 'L', 1386-1394, 'D', 1396-1404, 'E', 1406-1414, 'K', 1416-1424, 'G', 1426-1434, 'L', 1436-1444, 'D', 1446-1454, 'E', 1456-1464, 'K', 1466-1474, 'R', 1476-1484, 'L', 1486-1494, 'D', 1496-1504, 'E', 1506-1514, 'K', 1516-1524, 'G', 1526-1534, 'L', 1536-1544, 'D', 1546-1554, 'E', 1556-1564, 'K', 1566-1574, 'R', 1576-1584, 'L', 1586-1594, 'D', 1596-1604, 'E', 1606-1614, 'K', 1616-1624, 'G', 1626-1634, 'L', 1636-1644, 'D', 1646-1654, 'E', 1656-1664, 'K', 1666-1674, 'R', 1676-1684, 'L', 1686-1694, 'D', 1696-1704, 'E', 1706-1714, 'K', 1716-1724, 'G', 1726-1734, 'L', 1736-1744, 'D', 1746-1754, 'E', 1756-1764, 'K', 1766-1774, 'R', 1776-1784, 'L', 1786-1794, 'D', 1796-1804, 'E', 1806-1814, 'K', 1816-1824, 'G', 1826-1834, 'L', 1836-1844, 'D', 1846-1854, 'E', 1856-1864, 'K', 1866-1874, 'R', 1876-1884, 'L', 1886-1894, 'D', 1896-1904, 'E', 1906-1914, 'K', 1916-1924, 'G', 1926-1934, 'L', 1936-1944, 'D', 1946-1954, 'E', 1956-1964, 'K', 1966-1974, 'R', 1976-1984, 'L', 1986-1994, 'D', 1996-2004, 'E', 2006-2014, 'K', 2016-2024, 'G', 2026-2034, 'L', 2036-2044, 'D', 2046-2054, 'E', 2056-2064, 'K', 2066-2074, 'R', 2076-2084, 'L', 2086-2094, 'D', 2096-2104, 'E', 2106-2114, 'K', 2116-2124, 'G', 2126-2134, 'L', 2136-2144, 'D', 2146-2154, 'E', 2156-2164, 'K', 2166-2174, 'R', 2176-2184, 'L', 2186-2194, 'D', 2196-2204, 'E', 2206-2214, 'K', 2216-2224, 'G', 2226-2234, 'L', 2236-2244, 'D', 2246-2254, 'E', 2256-2264, 'K', 2266-2274, 'R', 2276-2284, 'L', 2286-2294, 'D', 2296-2304, 'E', 2306-2314, 'K', 2316-2324, 'G', 2326-2334, 'L', 2336-2344, 'D', 2346-2354, 'E', 2356-2364, 'K', 2366-2374, 'R', 2376-2384, 'L', 2386-2394, 'D', 2396-2404, 'E', 2406-2414, 'K', 2416-2424, 'G', 2426-2434, 'L', 2436-2444, 'D', 2446-2454, 'E', 2456-2464, 'K', 2466-2474, 'R', 2476-2484, 'L', 2486-2494, 'D', 2496-2504, 'E', 2506-2514, 'K', 2516-2524, 'G', 2526-2534, 'L', 2536-2544, 'D', 2546-2554, 'E', 2556-2564, 'K', 2566-2574, 'R', 2576-2584, 'L', 2586-2594, 'D', 2596-2604, 'E', 2606-2614, 'K', 2616-2624, 'G', 2626-2634, 'L', 2636-2644, 'D', 2646-2654, 'E', 2656-2664, 'K', 2666-2674, 'R', 2676-2684, 'L', 2686-2694, 'D', 2696-2704, 'E', 2706-2714, 'K', 2716-2724, 'G', 2726-2734, 'L', 2736-2744, 'D', 2746-2754, 'E', 2756-2764, 'K', 2766-2774, 'R', 2776-2784, 'L', 2786-2794, 'D', 2796-2804, 'E', 2806-2814, 'K', 2816-2824, 'G', 2826-2834, 'L', 2836-2844, 'D', 2846-2854, 'E', 2856-2864, 'K', 2866-2874, 'R', 2876-2884, 'L', 2886-2894, 'D', 2896-2904, 'E', 2906-2914, 'K', 2916-2924, 'G', 2926-2934, 'L', 2936-2944, 'D', 2946-2954, 'E', 2956-2964, 'K', 2966-2974, 'R', 2976-2984, 'L', 2986-2994, 'D', 2996-3004, 'E', 3006-3014, 'K', 3016-3024, 'G', 3026-3034, 'L', 3036-3044, 'D', 3046-3054, 'E', 3056-3064, 'K', 3066-3074, 'R', 3076-3084, 'L', 3086-3094, 'D', 3096-3104, 'E', 3106-3114, 'K', 3116-3124, 'G', 3126-3134, 'L', 3136-3144, 'D', 3146-3154, 'E', 3156-3164, 'K', 3166-3174, 'R', 3176-3184, 'L', 3186-3194, 'D', 3196-3204, 'E', 3206-3214, 'K', 3216-3224, 'G', 3226-3234, 'L', 3236-3244, 'D', 3246-3254, 'E', 3256-3264, 'K', 3266-3274, 'R', 3276-3284, 'L', 3286-3294, 'D', 3296-3304, 'E', 3306-3314, 'K', 3316-3324, 'G', 3326-3334, 'L', 3336-3344, 'D', 3346-3354, 'E', 3356-3364, 'K', 3366-3374, 'R', 3376-3384, 'L', 3386-3394, 'D', 3396-3404, 'E', 3406-3414, 'K', 3416-3424, 'G', 3426-3434, 'L', 3436-3444, 'D', 3446-3454, 'E', 3456-3464, 'K', 3466-3474, 'R', 3476-3484, 'L', 3486-3494, 'D', 3496-3504, 'E', 3506-3514, 'K', 3516-3524, 'G', 3526-3534, 'L', 3536-3544, 'D', 3546-3554, 'E', 3556-3564, 'K', 3566-3574, 'R', 3576-3584, 'L', 3586-3594, 'D', 3596-3604, 'E', 3606-3614, 'K', 3616-3624, 'G', 3626-3634, 'L', 3636-3644, 'D', 3646-3654, 'E', 3656-3664, 'K', 3666-3674, 'R', 3676-3684, 'L', 3686-3694, 'D', 3696-3704, 'E', 3706-3714, 'K', 3716-3724, 'G', 3726-3734, 'L', 3736-3744, 'D', 3746-3754, 'E', 3756-3764, 'K', 3766-3774, 'R', 3776-3784, 'L', 3786-3794, 'D', 3796-3804, 'E', 3806-3814, 'K', 3816-3824, 'G', 3826-3834, 'L', 3836-3844, 'D', 3846-3854, 'E', 3856-3864, 'K', 3866-3874, 'R', 3876-3884, 'L', 3886-3894, 'D', 3896-3904, 'E', 3906-3914, 'K', 3916-3924, 'G', 3926-3934, 'L', 3936-3944, 'D', 3946-3954, 'E', 3956-3964, 'K', 3966-3974, 'R', 3976-3984, 'L', 3986-3994, 'D', 3996-4004, 'E', 4006-4014, 'K', 4016-4024, 'G', 4026-4034, 'L', 4036-4044, 'D', 4046-4054, 'E', 4056-4064, 'K', 4066-4074, 'R', 4076-4084, 'L', 4086-4094, 'D', 4096-4104, 'E', 4106-4114, 'K', 4116-4124, 'G', 4126-4134, 'L', 4136-4144, 'D', 4146-4154, 'E', 4156-4164, 'K', 4166-4174, 'R', 4176-4184, 'L', 4186-4194, 'D', 4196-4204, 'E', 4206-4214, 'K', 4216-4224, 'G', 4226-4234, 'L', 4236-4244, 'D', 4246-4254, 'E', 4256-4264, 'K', 4266-4274, 'R', 4276-4284, 'L', 4286-4294, 'D', 4296-4304, 'E', 4306-4314, 'K', 4316-4324, 'G', 4326-4334, 'L', 4336-4344, 'D', 4346-4354, 'E', 4356-4364, 'K', 4366-4374, 'R', 4376-4384, 'L', 4386-4394, 'D', 4396-4404, 'E', 4406-4414, 'K', 4416-4424, 'G', 4426-4434, 'L', 4436-4444, 'D', 4446-4454, 'E', 4456-4464, 'K', 4466-4474, 'R', 4476-4484, 'L', 4486-4494, 'D', 4496-4504, 'E', 4506-4514, 'K', 4516-4524, 'G', 4526-4534, 'L', 4536-4544, 'D', 4546-4554, 'E', 4556-4564, 'K', 4566-4574, 'R', 4576-4584, 'L', 4586-4594, 'D', 4596-4604, 'E', 4606-4614, 'K', 4616-4624, 'G', 4626-4634, 'L', 4636-4644, 'D', 4646-4654, 'E', 4656-4664, 'K', 4666-4674, 'R', 4676-4684, 'L', 4686-4694, 'D', 4696-4704, 'E', 4706-4714, 'K', 4716-4724, 'G', 4726-4734, 'L', 4736-4744, 'D', 4746-4754, 'E', 4756-4764, 'K', 4766-4774, 'R', 4776-4784, 'L', 4786-4794, 'D', 4796-4804, 'E', 4806-4814, 'K', 4816-4824, 'G', 4826-4834, 'L', 4836-4844, 'D', 4846-4854, 'E', 4856-4864, 'K', 4866-4874, 'R', 4876-4884, 'L', 4886-4894, 'D', 4896-4904, 'E', 4906-4914, 'K', 4916-4924, 'G', 4926-4934, 'L', 4936-4944, 'D', 4946-4954, 'E', 4956-4964, 'K', 4966-4974, 'R', 4976-4984, 'L', 4986-4994, 'D', 4996-5004, 'E', 5006-5014, 'K', 5016-5024, 'G', 5026-5034, 'L', 5036-5044, 'D', 5046-5054, 'E', 5056-5064, 'K', 5066-5074, 'R', 5076-5084, 'L', 5086-5094, 'D', 5096-5104, 'E', 5106-5114, 'K', 5116-5124, 'G', 5126-5134, 'L', 5136-5144, 'D', 5146-5154, 'E', 5156-5164, 'K', 5166-5174, 'R', 5176-5184, 'L', 5186-5194, 'D', 5196-5204, 'E', 5206-5214, 'K', 5216-5224, 'G', 5226-5234, 'L', 5236-5244, 'D', 5246-5254, 'E', 5256-5264, 'K', 5266-5274, 'R', 5276-5284, 'L', 5286-5294, 'D', 5296-5304, 'E', 5306-5314, 'K', 5316-5324, 'G', 5326-5334, 'L', 5336-5344, 'D', 5346-5354, 'E', 5356-5364, 'K', 5366-5374, 'R', 5376-5384, 'L', 5386-5394, 'D', 5396-5404, 'E', 5406-5414, 'K', 5416-5424, 'G', 5426-5434, 'L', 5436-5444, 'D', 5446-5454, 'E', 5456-5464, 'K', 5466-5474, 'R', 5476-5484, 'L', 5486-5494, 'D', 5496-5504, 'E', 5506-5514, 'K', 5516-5524, 'G', 5526-5534, 'L', 5536-5544, 'D', 5546-5554, 'E', 5556-5564, 'K', 5566-5574, 'R', 5576-5584, 'L', 5586-5594, 'D', 5596-5604, 'E', 5606-5614, 'K', 5616-5624, 'G', 5626-5634, 'L', 5636-5644, 'D', 5646-5654, 'E', 5656-5664, 'K', 5666-5674, 'R', 5676-5684, 'L', 5686-5694, 'D', 5696-5704, 'E', 5706-5714, 'K', 5716-5724, 'G', 5726-5734, 'L', 5736-5744, 'D', 5746-5754, 'E', 5756-5764, 'K', 5766-5774, 'R', 5776-5784, 'L', 5786-5794, 'D', 5796-5804, 'E', 5806-5814, 'K', 5816-5824, 'G', 5826-5834, 'L', 5836-5844, 'D', 5846-5854, 'E', 5856-5864, 'K', 5866-5874, 'R', 5876-5884, 'L', 5886-5894, 'D', 5896-5904, 'E', 5906-5914, 'K', 5916-5924, 'G', 5926-5934, 'L', 5936-5944, 'D', 5946-5954, 'E', 5956-5964, 'K', 5966-5974, 'R', 5976-5984, 'L', 5986-5994, 'D', 5996-6004, 'E', 6006-6014, 'K', 6016-6024, 'G', 6026-6034, 'L', 6036-6044, 'D', 6046-6054, 'E', 6056-6064, 'K', 6066-6074, 'R', 6076-6084, 'L', 6086-6094, 'D', 6096-6104, 'E', 6106-6114, 'K', 6116-6124, 'G', 6126-6134, 'L', 6136-6144, 'D', 6146-6154, 'E', 6156-6164, 'K', 6166-6174, 'R', 6176-6184, 'L', 6186-6194, 'D', 6196-6204, 'E', 6206-6214, 'K', 6216-6224, 'G', 6226-6234, 'L', 6236-6244, 'D', 6246-6254, 'E', 6256-6264, 'K', 6266-6274, 'R', 6276-6284, 'L', 6286-6294, 'D', 6296-6304, 'E', 6306-6314, 'K', 6316-6324, 'G', 6326-6334, 'L', 6336-6344, 'D', 6346-6354, 'E', 6356-6364, 'K', 6366-6374, 'R', 6376-6384, 'L', 6386-6394, 'D', 6396-6404, 'E', 6406-6414, 'K', 6416-6424, 'G', 6426-6434, 'L', 6436-6444, 'D', 6446-6454, 'E', 6456-6464, 'K', 6466-6474, 'R', 6476-6484, 'L', 6486-6494, 'D', 6496-6504, 'E', 6506-6514, 'K', 6516-6524, 'G', 6526-6534, 'L', 6536-6544, 'D', 6546-6554, 'E', 6556-6564, 'K', 6566-6574, 'R', 6576-6584, 'L', 6586-6594, 'D',

A;Contents: annotation; X-ray crystallography, 1.9 angstroms  
C;Comment: This protein is excited by the photoprotein seagurin (see PIR:AQJFNV) emitting  
C;Comment: The chromophore of this protein is formed by modification of Ser-dehydro-tyr-  
C;Genetics:  
A;Gene: GPP  
A;Introns: 69/3; 167/3  
C;Superfamily: green-fluorescent protein  
C;Keywords: chromoprotein; luminescence  
F;65-67/Cross-link: 5-imidazolinone (Ser-Gly) #status experimental  
F;66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match 97.4%; Score 1241; DB 1; Length 238;  
Best Local Similarity 97.1%; Pred. No. 1.2e-36;  
Matches 231; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVELDGVNGHKFSVSGEGEDATYKGLTKFKICTTGKLPVWPPTL 61  
DB 1 MSKGEELFTGVVPIVLVELDGVNGHKFSVSGEGEDATYKGLTKFKICTTGKLPVWPPTL 60

QY 62 VTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121  
DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120

QY 122 NRIELAGIDPKEDGNILGHKLEYNVSHVYIMADKQNGIKVNFIRHNIEDGSVOLAD 181  
DB 121 NRIELAGIDPKEDGNILGHKLEYNVSHVYIMADKQNGIKVNFIRHNIEDGSVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMHVLGFTVTAAGITLGMDELYK 239  
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMHVLGFTVTAAGITLGMDELYK 238

RESULT 2  
H72228  
hypothetical protein TM1624 - Thermotoga maritima (strain MSB8)  
C;Species: Thermotoga maritima  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C;Accession: H72228  
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.

Nature 399, 323-329, 1999  
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A;Reference number: A72200; MUID:99287316; PMID:10360571  
A;Accession: H72228  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-785 <ARN>  
A;Cross-references: GB:AE001806; GB:AE000512; NID:g4982196; PIDN:AAD36691.1; PID:g498219  
A;Experimental source: strain MSB8  
C;Genetics:  
A;Gene: TM1624

Query Match 8.2%; Score 105; DB 2; Length 785;  
Best Local Similarity 19.7%; Pred. No. 0.88;  
Matches 46; Conservative 32; Mismatches 71; Indels 84; Gaps 7;

QY 3 SKGEELFTGVVPIVLVELDGVNGHKFSVSGEGEDATYKGLTKFKICTTGKLPVWPPTLV 62  
DB 15 NEGRFSPEGTVPVQVQAD-----LVRKGLLPHPYVGM- 46

QY 63 TTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLVN 122  
DB 47 -----NEDLFEIEDRWIYVEREFKEDVKEGERDLVVEGVDTLN 88

QY 123 RIELKGIDFKEDGNILGHKLEYNVSHVYIMADKQNGIKVNFIRHNIEDGSVOLADH 182  
DB 89 DVYLVGYL---GSTEDMFIEYRFDVTNLV---KEKNHLKVYIK-----SPIRVPKT 134

QY 183 YQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMHVLGFTVTAAGITLGM 235  
DB 135 LEQNYGLVGGP-----BDP-----IRGYRKAQSYGWD 163

RESULT 3

H64102  
leucine-tRNA ligase (EC 6.1.1.4) - Haemophilus influenzae (strain Rd KW20)

N;Alternate names: leucyl-tRNA synthetase

C;Species: Haemophilus influenzae

C;Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 03-Jun-2002

C;Accession: H64102

R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrman, J.L.; Geoghagen, N.S.M.

Science 269, 496-512, 1995

A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.

A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A;Reference number: A64000; MUID:95350630; PMID:7542800

A;Accession: H64102

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-861 <TIGR>

A;Cross-references: GB:U32774; GB:L42023; NID:g1573942; PIDN:AAC22581.1; PID:g1573943; T;

C;Genetics:

C;Superfamily: leucine-tRNA ligase

C;Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 7.2%; Score 91.5; DB 2; Length 861;  
Best Local Similarity 24.1%; Pred. No. 13;  
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;

QY 50 TTGKLPVWPPTLVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFK 103

DB 314 TGDKLPI-WVANFVLMHYGTGAVMAVPAH-DQRDF-----EFAQKYSLPINQVIAPLA 364

QY 104 DGNKYTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNVSHVYIMADK-OKNGI 162

DB 365 DEEIDLTKQAFVEHGHKLVNSDFDGKNF--DGAFG-----IADKLEKLG 408

QY 163 ---KVNFKIRH-----NIEDGSVOLADHYQQYQNTPIGDGPVLLPDNHYL- 202

DB 409 GKRQVNYRLRDWGVSRQRYWGAPIPMLTLENGDVVPA-----PMEDLPILLPEDVWMD 461

QY 203 STQSALSKDPN 213

DB 462 GVKSPINADPN 472

RESULT 4

T06586

DNA-binding protein PD2 - garden pea

C;Species: Pisum sativum (garden pea)

C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 08-Oct-1999

C;Accession: T06586

R;Sato, N.; Kazuno, A.A.; Ohta, N.; Ohshima, K.

submitted to the EMBL Data Library, June 1996

A;Description: Identification of a novel family of DNA-binding proteins with two AT-hook

A;Reference number: Z15774

A;Accession: T06586

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-632 <SAT>

A;Cross-references: EMBL:X98740; NID:e995229; PIDN:CAA67292.1; PID:e275185

A;Experimental source: cv. Alaska

Query Match 7.1%; Score 90; DB 2; Length 632;  
Best Local Similarity 23.3%; Pred. No. 12;  
Matches 49; Conservative 26; Mismatches 79; Indels 56; Gaps 7;

QY 16 LVLEDGVDVNGHKFSVSGEGEDATYKGLTKLTKFKICTTGKLPVWPPTLVTTLSYGVQCFSR 75

DB 363 IVQGRDVGSKVDVINKESNEATIPENK----PTBPCLDVEQELAAATTM----- 408

QY 76 PDHMKQHDFFKSAMPEGYV-----QERTIFFKDDGNKYKTRAEVKFEGDTLVNRIE 125



Db 409 PPSAKNVNLTDLIVFTFLRSVARTSGREGSEELKDSNLSRLRTKKLELEQGNKS-B 467

QY 126 LKIGIDFKEDGNILGHKLE-----NYNSHNYYIMADKQK-----N 160

Db 468 LKGIETDNTSLDDEFENALGNKILKEISNPRHDVESHNSHTNKQVTVSHQKAETNN 527

QY 161 GIKVNFKIRHNIEDG-----SVOLADHYQ 184

Db 528 QSQVEDVAKNIQDDSKPSESLHKADKYR 557

RESULT 5

E82590

leucyl-tRNA synthetase XF2176 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000

C:Accession: E82590

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717; PMID:10910347

A>Note: for a complete list of authors see reference number A59328 below

A:Accession: E82590

A>Status: preliminary

A:Molecule type: DNA

C:Genetics:

A:Cross-references: GB:AE004031; GB:AE003849; NID:g9107309; PIDN:AAF84975.1; GSPDB:GN001

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H

as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laig

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.B.; Marques, M.V.; Martins, E

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, R.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF2176

C:Superfamily: leucine-tRNA ligase

Query Match 7.0%; Score 89.5; DB 2; Length 887;

Best Local Similarity 22.2%; Pred. No. 21;

Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;

QY 50 TTCKLPVPWPTLVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNY-- 107

Db 329 TNEQLPV-WVANFVLMAYGTGAVMVGPHDQDQEF--ANKYGLPRTQVIALKEPKNQDE 385

QY 108 -----KTRAEVKFEGDTLVNRIELKGDIFDKEDGNILGHKLEYNVNSHNVI 153

Db 386 STWEPDVRWDYADKTR---EFE---LINSAPFDGLDYQDAFVLAERFE----- 429

QY 154 MADKQKNG-IKVNFKIRHNIEDGSVOLADHYQOQNTPI-----GGGPVLLPDN 199

Db 430 ---RQGRQRVNNYLR-----DWGVSQRVWGCPVPVYCTGCAVPVPEDQLPVILPEN 482

QY 200 -HYLSQTQSALSQDPNEKR 216

Db 483 VAFSGTGSPIKTPDPEWRK 500

RESULT 6

D83917

DNA topoisomerase IV subunit B BH2140 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C:Accession: D83917

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hiran

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: D83917

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-655 <STO>

A:Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BAB05859.1; GSPDB:GN001

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH2140

C:Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B

Query Match 6.9%; Score 88.5; DB 2; Length 655;

Best Local Similarity 21.9%; Pred. No. 17;

Matches 53; Conservative 37; Mismatches 95; Indels 57; Gaps 9;

QY 22 DVNGHK---FSVSGEGEGDAT---YGLTKLTKPI-----CTTGKLPVPWP 59

Db 63 NVTIHKQSVSRDEGRGMPTGMHKLKPTPEVILTVLHAGGKFGQGGYATSGLHGVA 122

QY 60 TLVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQER-----TIFEKDDG----- 105

Db 123 SYVNALSEWLIVEIKRDGVVYQRFENGKPKSTTLEKKGKTRQGTGTHFKPDPTVFSST 182

QY 106 --NYKTRAEVKFEGDTLVNRIELKGDIFDKEDGNILGHKLEYNVNSHNVIADK----- 157

Db 183 NFNVTLSRLREAAFLKGLKIELVDLRDDYKEVFH--YEDGKAFVYLNEDKEITLHPV 241

QY 158 -----QKNGIKVNFKIRHNIEDGSVOLADHYQOQNTPIGDPVLLPDNHYLSQTQSALSQDP 212

Db 242 VFFNGESNGIEIEFAFOFN--DGYTENVLFSVNNVVRTKD-----GTHELGAKTAMTRAV 294

QY 213 NE 214

Db 295 NE 296

RESULT 7

E70390

iron-sulfur cofactor synthesis protein nifs - Aquifex aeolicus

N:Contains: L-cysteine sulfurtransferase (EC 2.8.1.-)

C:Species: Aquifex aeolicus

C>Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 07-Dec-1999

C:Accession: E70390

R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove

V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: E70390

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-370 <AQF>

A:Cross-references: GB:AE000720; NID:g2983529; PIDN:AAC07111.1; PID:g2983536; GB:AE00065

A:Experimental source: strain VF5

C:Genetics:

A:Gene: nifs1

C:Superfamily: nitrogen fixation protein nifs

C:Keywords: phosphoprotein; pyridoxal phosphate; sulfurtransferase

F:195/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

F:318/Active site: Cys (cysteine persulfide intermediate) #status predicted

Query Match 6.9%; Score 87.5; DB 2; Length 370;

Best Local Similarity 25.4%; Pred. No. 9.7;

Matches 49; Conservative 30; Mismatches 91; Indels 23; Gaps 8;

QY 4 KGEELFTGVV---PILVELD---GDVNGHKF-SVSGEG-----EGDATYKLTIAFICT 50

Db 164 KGVPVLTDAVQAIKPIELKINISATFSGHKFPAIKGSGFLYISDEANYEPLVGGQGE 223

QY 51 TGKLP-----VWPVTLVTTLSYGVQCFSRYPDHMKQ-HDFFKSAMPEGYVQERTIFFKOD 104

Db 224 NGRSGTENVVGLSLAKALEIIVFSRYQEQKLRDLFENLLLEA-LPDAQIVGKA 282  
 QY 105 GNYKTRAEV---KFGEDTLVNRIELKGDIFKEDGNILGHKLEYNHSHVYIMADKQKNG 161  
 Db 283 ERSFSSISSVIMPRFFGAEIIVNKLSEGIYGTSGACLSGGEYEPNKLKMGFQERKALRM 342  
 QY 162 IKVNFKIRHNIED 174  
 Db 343 VRFSGLLNKEEE 355

RESULT 8  
 AC0582  
 leucyl-tRNA synthetase [imported] - Salmonella enterica subsp. enterica serovar Typhi (S  
 C:Species: Salmonella enterica subsp. enterica serovar Typhi  
 A:Note: this species has also been called Salmonella typhi  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
 C:Accession: AC0582  
 R;Farkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
 th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
 , S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
 A:Reference number: AB0502; MUID:21534947; PMID:11677608  
 A:Accession: AC0582  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-860 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD05125.1; PID:g16501899; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY0699  
 C:Superfamily: leucine-tRNA ligase

Query Match 6.9%; Score 87.5; DB 2; Length 860;  
 Best Local Similarity 23.3%; Pred. No. 29;  
 Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;

QY 50 TTGKLPVMPPTLVTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNYKT 109  
 Db 314 TGEIIPV-WAANFVLMVEYGTGAVMAVFGH-DQRD-YEFASKYGLTIKPVILAADSGEPDL 370  
 QY 110 RAEVKPEGDTLVNRIELKGDIFKEDGNILGHKLEYNHSHVYIMADKQKNGKYNFKIR 169  
 Db 371 SEQALTEKGVLFNSGDFGLAFAPAFNAIADKL-----AEKVGGERKYNRLR 418  
 QY 170 H-----NIEDGSQLADHYQQTPTIGDGPVLLPDNHYL-STQSALSKDP 212  
 Db 419 DWGVSQRQRYGAPIPVLTLEDGTV-----LPTPDQLPVLIPEDVMDGITSPIKADP 471

RESULT 9  
 A36028  
 DNA-directed DNA polymerase (EC 2.7.7.7) II catalytic chain - yeast (Saccharomyces cerev  
 N;Alternate names: DNA-directed DNA polymerase II chain A; protein N0825; protein YNL262  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
 C:Accession: A36028; B36028; S60919; S63235; S65121  
 R;Morrison, A.; Araki, H.; Clark, A.B.; Hamatake, R.K.; Sugino, A.  
 Cell 62, 1143-1151, 1990  
 A:Title: A third essential DNA polymerase in Saccharomyces cerevisiae.  
 A:Reference number: A36028; MUID:90381771; PMID:2169349  
 A:Accession: A36028  
 A:Molecule type: DNA  
 A:Residues: 1-2222 <MOR>  
 A:Cross-references: GB:M60416; GB:M36724; NID:g171408; PIDN:AA88711.1; PID:g171409  
 A:Accession: B36028  
 A:Molecule type: protein  
 A:Residues: 1214-1216, 'X', 1218-1221 <MO2>  
 R;Sen-Gupta, M.; Lyck, R.; Fleig, U.; Niedenthal, R.N.; Hegemann, J.M.  
 submitted to the EMBL Data Library, October 1995  
 A:Description: The sequence of a 24152 bp segment from the left arm of chromosome XIV fd

A:Reference number: S60909  
 A:Accession: S60919  
 A:Molecule type: DNA  
 A:Residues: 1-2221 <SEN>  
 A:Cross-references: EMBL:X92494; NID:g1045236; PIDN:CAA63235.1; PID:g1045247  
 R;Sen-Gupta, M.; Lyck, R.; Niedenthal, R.K.; Fleig, U.N.; Hegemann, J.H.  
 submitted to the Protein Sequence Database, April 1996  
 A:Reference number: S63235  
 A:Accession: S63235  
 A:Molecule type: DNA  
 A:Residues: 1-2222 <SEW>  
 A:Cross-references: EMBL:X92494; NID:g1302316; PIDN:CAA96169.1; PID:g1302317; GSPDB:GN001  
 A:Experimental source: strain S288C  
 R;Sen-Gupta, M.; Lyck, R.; Fleig, U.; Niedenthal, R.K.; Hegemann, J.H.  
 yeast 12, 50S-514, 1996  
 A:Title: The sequence of a 24 152 bp segment from the left arm of chromosome XIV from Sac  
 A:Reference number: S65111; MUID:96310631; PMID:8740425  
 A:Accession: S65121  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-2221 <SEF>  
 A:Cross-references: EMBL:X92494; NID:g1045236; PIDN:CAA63235.1; PID:g1045247  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995  
 C:Genetics:  
 A:Gene: SGD:POL2; DUN2: MIPS:YNL262w  
 A:Cross-references: SGD:S0005206; MIPS:YNL262w  
 A:Map position: 14L  
 C:Superfamily: DNA-directed DNA polymerase II  
 C:Keywords: DNA binding; nucleotidyltransferase; nucleus; zinc finger

Query Match 6.9%; Score 87.5; DB 1; Length 2222;  
 Best Local Similarity 28.2%; Pred. No. 1e+02;  
 Matches 37; Conservative 14; Mismatches 49; Indels 31; Gaps 7;

QY 54 LPVPWP-TLVTTLSYGVCFSRYPDH-----KQHDFFKSAMPEGYV-----QERTI 99  
 Db 883 LPKSPETVFTTLENGKLYLSYPCSMNVVRVHQKFTNHQYQELKDPPLNVYETHSENTI 942  
 QY 100 FKDDGNYKTR--AEVKFEGDTLVNR-----IELKGIDFKEDGNILGHKLEYNYN 147  
 Db 943 FFEVDGPKAMILPSSKEEGKIKRYAVFNEDGSLAELKGFELKRGEL---OLIKNFQ 999  
 QY 148 S--HNVYIMAD 156  
 Db 1000 SDIFKVFLEGD 1010

RESULT 10  
 D71614  
 hypothetical protein PFB0460c - malaria parasite (Plasmodium falciparum)  
 C:Species: Plasmodium falciparum  
 C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
 C:Accession: D71614  
 R;Gardner, M.J.; Tetteilin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
 ; Perlea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;  
 Science 282, 1126-1132, 1998  
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
 A:Reference number: A71600; MUID:99021743; PMID:9804551  
 A:Accession: D71614  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-2573 <GAP>  
 A:Cross-references: GB:AE001396; GB:AE001362; NID:g3845188; PIDN:AACT71881.1; PID:g3845190  
 A:Experimental source: clone 3D7  
 C:Genetics:  
 A:Gene: PFB0460c

Query Match 6.9%; Score 87.5; DB 2; Length 2573;  
 Best Local Similarity 26.2%; Pred. No. 1.2e+02;  
 Matches 34; Conservative 30; Mismatches 53; Indels 13; Gaps 5;

QY 94 VOERTIFFKD--DGNKYTRAEVKFEGDTLVNRIELKGDIFKEDGNILGHKLEYN--YN 149  
 ::::: ||||| :|: ||| |||: :|: |||



QY	197	P--DNHYLSTQALSQDPNEKDHVMVLLGFTVTAAGITLG	233
Db	594	PGSDNKYYKLSADVQGFYPLDRDHLWVWSAKASAGYANG	632
RESULT 14			
F64102			
protective surface antigen D-15 - Haemophilus influenzae (strain Rd KW20)			
C:Species: Haemophilus influenzae			
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 18-Sep-1998			
C:Accession: F64102			
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M. Science 269, 496-512, 1995			
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.			
A:Reference number: A64000; MUID:95350630; PMID:7542800			
A:Accession: F64102			
A>Status: nucleic acid sequence not shown; translation not shown			
A:Molecule type: DNA			
A:Residues: 1-808 <TIGR>			
A:Cross-references: GB:L42023; TIGR:HI0917			
C:Superfamily: protective surface antigen D-15			
C:Keywords: surface antigen			
Query Match 6.8%; Score 86.5; DB 2; Length 808;			
Best Local Similarity 21.9%; Pred. No. 33;			
Matches 48; Conservative 30; Mismatches 78; Indels 63; Gaps 11;			
QY	65	LSGVQCFSRYPDHMKOHDF-----FKSAMPEGYVQE-----RTI 99	
Db	440	IGYTESGHSYQASVKODNPLGTGAAVSAGTAKNDYGVNLGYTEPYTKDGVSLGGNV 499	
QY	100	FPKDDGNYKTRABVKFEGDTLVNRIELKGIDFKEDGNI---LGH-----KLEYNYS 148	
Db	500	FFENYDNSKSDTSSNYKRTYGSNVTL-GFPVNNNSYVGLGHTYTKISNPALEYN--- 555	
QY	149	HNYYIMADKQK-NGIKVFKIRINIEDGVSQVLADHYQQ-----NTPIGDGPVLL 196	
Db	556	RNLYIQSMKFKNGIKITN-----DFDFSGWYNSLNRGYFPTKGVKASIG-GRVTI 606	
QY	197	P--DNHYLSTQALSQDPNEKDHVMVLLGFTVTAAGITLG	233
Db	607	PGSDNKYYKLSADVQGFYPLDRDHLWVWSAKASAGYANG	645
RESULT 15			
G81355			
tRNA (uracil-5-)-methyltransferase (EC 2.1.1.35) Cj0831c [imported] - Campylobacter jejuni			
C:Species: Campylobacter jejuni			
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002			
C:Accession: G81355			
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillingworth, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrett, T.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrett, T. Nature 403, 665-668, 2000			
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp			
A:Reference number: A81250; MUID:20150912; PMID:10688204			
A:Accession: G81355			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-357 <PAR>			
A:Cross-references: GB:A1139076; GB:A1111168; NID:g968128; PIDN:CAB73096.1; PID:g696827			
A:Experimental source: serotype O2, strain NCTC 11168			
C:Genetics:			
A:Gene: trnA; Cj0831c			
C:Keywords: methyltransferase; S-adenosylmethionine			
Query Match 6.8%; Score 86; DB 2; Length 357;			
Best Local Similarity 24.8%; Pred. No. 12;			
Matches 30; Conservative 18; Mismatches 39; Indels 34; Gaps 5;			
QY	80	KQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKF--EGDTLV-----NRIELKG 128	

Db	14	EKHSFIKKYFKEFYTKDFKLFASKDKHYRTRAELSFYHENDTLFYAMFDPKSKKKYIIEY	73
QY	129	IDFKED-----GNILGHKLEYNSHNYYIMADKQKNGIKVNFKIRHNIE	173
Db	74	LDFADEKICAFMPRLLEYLRQDNKLEKL-----FGVEFLTTRQEE--LSITLLYHKNIE	125
QY	174	D 174	
Db	126	D 126	

Search completed: June 21, 2004, 16:01:57  
Job time : 10.3333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:42:24 ; Search time 6.44444 Seconds  
(without alignments)  
1931.085 Million cell updates/sec

Title: US-09-887-784-2221  
Perfect score: 1274  
Sequence: 1 MYSKGELFTGVVPILVELD.....VLlGFVTAAGITLGMDELK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1247	97.9	238	1	GFP_AEQVI
2	91.5	7.2	861	1	SYL_HAEIN
3	89.5	7.0	879	1	SYL_XYLFA
4	87.5	6.9	860	1	SYL_SALTI
5	87.5	6.9	860	1	SYL_SALTY
6	87.5	6.9	2222	1	DPOE_YEAST
7	87	6.8	689	1	AC2L_HUMAN
8	87	6.8	874	1	SLAP_BACLI
9	86.5	6.8	533	1	CP51_CANGA
10	86.5	6.8	795	1	D152_HAEIN
11	86.5	6.8	797	1	D151_HAEIN
12	86.5	6.8	879	1	SYL_XYLFT
13	86.5	6.8	1603	1	VIT4_CABEL
14	86	6.8	357	1	TRMA_CAMJE
15	85.5	6.7	788	1	DPOL_HPBHE
16	85.5	6.7	793	1	D153_HAEIN
17	85.5	6.7	886	1	ITH3_MESAU
18	85.5	6.7	941	1	GUN_BACS6
19	85	6.7	439	1	SY62_DISOM
20	84.5	6.6	501	1	AMPA_WICBR
21	84.5	6.6	613	1	PEPF_WYCPU
22	84.5	6.6	859	1	SYL_SHEON
23	84	6.6	353	1	HIS7_BUCAI
24	84	6.6	366	1	SET7_HUMAN
25	84	6.6	504	1	YC03_KLEPN
26	83.5	6.6	538	1	GRBE_RAT
27	83.5	6.6	658	1	ADAS_HUMAN
28	83.5	6.6	1164	1	BAG_STRAG
29	83	6.5	461	1	PSBC_CTAPA
30	83	6.5	774	1	AMY2_SCHPO
31	82.5	6.5	533	1	NIFD_CLOPA
32	82	6.4	682	1	PRC_ECOLI
33	82	6.4	752	1	NECI_RAT

34	81.5	6.4	589	1	SYD_HAEDU
35	81	6.4	336	1	YD48_METJA
36	81	6.4	682	1	AC2L_MOUSE
37	81	6.4	1224	1	COPA_HUMAN
38	80.5	6.3	393	1	TRME_HELPY
39	80.5	6.3	658	1	ADAS_CAVPO
40	80.5	6.3	860	1	SYL_ECOS7
41	80.5	6.3	860	1	SYL_ECOL6
42	80.5	6.3	860	1	SYL_ECOLI
43	80	6.3	461	1	PSBC_CHLEU
44	80	6.3	737	1	OPT1_DROME
45	79.5	6.2	312	1	TRXB_CHDMU

ALIGNMENTS

RESULT 1					
GFP_AEQVI					
ID	GFP_AEQVI	STANDARD;	PRT;	238 AA.	
AC	P42212; Q17104; Q27903;				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Green fluorescent protein.				
GN	GFP.				
OS	Aequorea victoria (Jellyfish).				
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;				
OC	Aequoreidae; Aequorea.				
OX	NCBI_TaxID=6100;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RX	MEDLINE=92175527; PubMed=1347277;				
RA	Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,				
RA	Cormier M.J.;				
RT	"Primary structure of the Aequorea victoria green-fluorescent				
RT	protein.";				
RL	Gene 111:229-233(1992).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=94185810; PubMed=8137953;				
RA	Inouye S., Tsuji F.I.;				
RT	"Aequorea green fluorescent protein. Expression of the gene and				
RT	fluorescence characteristics of the recombinant protein.";				
RL	FEBS Lett. 341:277-280(1994).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=97299832; PubMed=9154981;				
RA	Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;				
RT	"Enhanced expression in tobacco of the gene encoding green fluorescent				
RT	protein by modification of its codon usage.";				
RL	Plant Mol. Biol. 33:989-999(1997).				
RN	[4]				
RP	CHROMOPHORE.				
RX	MEDLINE=93192221; PubMed=8448132;				
RA	Cody C.W., Praher D.C., Westler W.M., Prendergast F.G., Ward W.W.;				
RT	"Chemical structure of the hexapeptide chromophore of the Aequorea				
RT	green-fluorescent protein.";				
RL	Biochemistry 32:1212-1218(1993).				
RN	[5]				
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).				
RX	MEDLINE=96355665; PubMed=8703075;				
RA	Ormoe M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,				
RA	Remington S.J.;				
RT	"Crystal structure of the Aequorea victoria green fluorescent				
RT	protein.";				
RL	Science 273:1392-1395(1996).				
RN	[6]				
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).				
RX	MEDLINE=98294543; PubMed=9631087;				
RA	Yang F., Moss L.G., Phillips G.N. Jr.;				
RT	"The molecular structure of green fluorescent protein.";				
RL	Nat. Biotechnol. 14:1246-1251(1996).				

Q7vnf0	haemophilus
Q58743	methanococc
Q99nb1	mus musculus
P53621	homo sapien
Q25443	helicobacte
P97275	cavia porce
Q8xbn8	escherichia
Q8fjy9	escherichia
P07813	escherichia
Q08684	chlamydomon
P91679	drosophila
Q9pck7	chlamydia m

[7]  
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.  
 RX MEDLINE=98455509; PubMed=9782051;  
 RA Wachter R.M., Eislinger M.A., Kallio K., Hanson G.T., Remington S.J.;  
 RT "Structural basis of spectral shifts in the yellow-emission variants  
 of green fluorescent protein."; Structure 6:1267-1277(1998).  
 RL [8]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RX MEDLINE=99238303; PubMed=10220315;  
 RA Eislinger M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J.;  
 RT "Structural and spectral response of green fluorescent protein  
 variants to changes in pH."; Biochemistry 38:5296-5301(1999).  
 RL Biochemistry 38:5296-5301(1999).  
 CC -1- FUNCTION: Energy-transfer acceptor. Its role is to transduce the  
 blue chemiluminescence of the protein aequorin into green  
 fluorescent light by energy transfer. Fluoresces in vivo upon  
 receiving energy from the Ca(2+)-activated photoprotein aequorin.  
 CC Absorbs light maximally at 395 nm and exhibits a smaller  
 absorbance peak at 470 nm. The fluorescence emission spectrum  
 peaks at 509 nm with a shoulder at 540 nm.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- TISSUE SPECIFICITY: Photocytes.  
 CC -1- PTM: Contains a covalently attached chromophore, which is composed  
 of modified amino acid residues. The chromophore is formed upon  
 cyclization of the residues Ser-dehydroTyr-Gly.  
 CC -1- BIOTECHNOLOGY: Has become a useful and ubiquitous tool for making  
 chimeric proteins of GFP linked to other proteins where it  
 functions as a fluorescent protein tag. GFP tolerates N- and C-  
 terminal fusion to a broad variety of proteins. It has been  
 expressed in bacteria, yeast, slime mold, plants, Drosophila,  
 zebrafish, and in mammalian cells. As a noninvasive fluorescent  
 marker in living cells, it allows for a wide range of applications  
 where it may function as a cell lineage tracer, reporter of gene  
 expression, or as a measure of protein-protein interactions.  
 CC -1- DATABASE: NAME-Protein Spotlight;  
 CC NOTE=Issue 11 of June 2001;  
 CC WWW="http://www.expasy.org/spotlight/articles/sptl-011.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M62654; AAA27722.1; -;  
 DR EMBL; M62653; AAA27721.1; -;  
 DR EMBL; L29345; AAA58246.1; -;  
 DR EMBL; X96418; CAA65278.1; -;  
 DR FIR; JS0692; JQ1514.  
 DR PDB; 1B9C; 17-NOV-00.  
 DR PDB; 1BFP; 07-JUL-97.  
 DR PDB; 1C4F; 14-JUN-00.  
 DR PDB; 1EWA; 08-NOV-96.  
 DR PDB; 1EMB; 16-JUN-97.  
 DR PDB; 1EMC; 20-AUG-97.  
 DR PDB; 1EME; 20-AUG-97.  
 DR PDB; 1EMF; 20-AUG-97.  
 DR PDB; 1EMG; 12-MAY-99.  
 DR PDB; 1EMK; 20-AUG-97.  
 DR PDB; 1EML; 20-AUG-97.  
 DR PDB; 1EMW; 20-AUG-97.  
 DR PDB; 1F09; 17-NOV-00.  
 DR PDB; 1F0B; 17-NOV-00.  
 DR PDB; 1GFL; 11-JAN-97.  
 DR PDB; 1HGT; 15-JAN-02.  
 DR PDB; 1HUY; 04-JUL-01.  
 DR PDB; 1JBY; 07-JAN-03.  
 DR PDB; 1JBJ; 07-JAN-03.  
 DR PDB; 1KP5; 28-AUG-02.  
 DR PDB; 1KYP; 10-APR-02.

DR PDB; 1KYR; 10-APR-02.  
 DR PDB; 1KYS; 10-APR-02.  
 DR PDB; 1YFP; 28-OCT-98.  
 DR PDB; 2EMD; 20-AUG-97.  
 DR PDB; 2EMN; 20-AUG-97.  
 DR PDB; 2EMO; 20-AUG-97.  
 DR InterPro; IPR000786; GFP like.  
 DR InterPro; IPR009017; GFP like.  
 DR Pfam; PF01353; GFP, 1.  
 DR PRINTS; PD01229; GFP, 1.  
 DR PRINTS; PD013756; Green fl protein; 1.  
 KW Luminescence; 3D-structure.  
 FT CROSSLINK 65 67 5-imidazolinone (Ser-Gly).  
 FT MOD RES 66 66 2,3-DIDEHYDROTYROSINE.  
 FT VARIANT 100 100 F -> Y.  
 FT VARIANT 108 108 T -> S.  
 FT VARIANT 141 141 L -> M.  
 FT VARIANT 219 219 V -> I.  
 FT CONFLICT 2 2 S -> G (IN REF. 3).  
 FT CONFLICT 25 25 H -> Q (IN REF. 2).  
 FT CONFLICT 80 80 Q -> R (IN REF. 3).  
 FT CONFLICT 157 157 Q -> P (IN REF. 2).  
 FT CONFLICT 172 172 E -> K (IN REF. 2).  
 FT HELIX 4 8  
 FT STRAND 12 22  
 FT TURN 23 24  
 FT STRAND 25 36  
 FT TURN 37 40  
 FT STRAND 41 48  
 FT TURN 49 50  
 FT HELIX 57 60  
 FT TURN 61 63  
 FT HELIX 69 71  
 FT STRAND 73 73  
 FT HELIX 76 81  
 FT HELIX 83 86  
 FT TURN 87 90  
 FT STRAND 92 100  
 FT TURN 101 102  
 FT STRAND 105 115  
 FT TURN 116 117  
 FT STRAND 118 128  
 FT TURN 132 133  
 FT TURN 135 139  
 FT STRAND 141 141  
 FT STRAND 148 155  
 FT TURN 156 159  
 FT STRAND 160 171  
 FT TURN 172 173  
 FT STRAND 176 187  
 FT STRAND 199 208  
 FT TURN 211 212  
 FT STRAND 217 227  
 SQ SEQUENCE 238 AA; 26886 MW; EA5A6F21FBFB6E05 CRC64;  
 Query Match 97.9%; Score 1247; DB 1; Length 238;  
 Best Local Similarity 98.3%; Pred. No. 4.7e-96;  
 Matches 234; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYKLTFLKFICTTGKLPVWPPTL 61  
 :|||||  
 Db 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYKLTFLKFICTTGKLPVWPPTL 60  
 QY 62 VTTLISYGVQCFRSYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNKYKTRAEVKFEGDTLV 121  
 :|||||  
 Db 61 VTTFISYGVQCFRSYPDHMKQHDFFPKSAMPEGVYQERTIFFKDDGNKYKTRAEVKFEGDTLV 120  
 QY 122 NRIELKGIDFKEDGNILGHKLRYKINSHYNYIMADKKNGIKVNFKIRHNIEDGSVQLAD 181  
 :|||||  
 Db 121 NRIELKGIDFKEDGNILGHKLRYKINSHYNYIMADKKNGIKVNFKIRHNIEDGSVQLAD 180  
 QY 182 HYQONTPIGDPVLLPDNHYLTQSALS KDPNEKRDMVLLGFVTAAGITLGMDELYK 239  
 :|||||

```
Db 181 HYQONTPIGDPVLLPDNHYLSQTSALSQKDPNEKRDHMLLEFVTAAGITHGMDELYK 238
RESULT 2
SYL_HAEIN
ID SYL_HAEIN STANDARD; PRT; 861 AA.
AC P43827;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--trNA ligase) (LeuRS).
GN LEUS OR H10921.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-P., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32774; AAC22581.1; -
CC PIR; H64102; H64102.
CC TIGR; H10921; -
CC HAMAP; MF_00049; -; 1.
CC InterPro; IPR002302; Leu-trNASynt1a.
CC InterPro; IPR002300; tRNA-synt_1a.
CC InterPro; IPR001412; tRNA-synt_1.
CC InterPro; IPR009008; ValRS_f1ers_edit.
CC Pfam; PF00133; tRNA-synt 1; 1.
CC PRINTS; PR00985; TRNASYNTHLEU.
CC TIGRFAMs; TIGR00396; leuS_bact; 1.
CC PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY)
SQ SEQUENCE 861 AA; 97750 MW; EB93304F6B4C8FB7 CRC64;
Query Match 7.2%; Score 91.5; DB 1; Length 861;
Best Local Similarity 24.1%; Pred. No. 6.5;
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;
QY 50 TTGKLPVNPPTLVTTLISYGVQCFSRYPDHMKQDFFKSAWPEGYQVETIFFKD----- 103
Db 314 TGDKLP1-WVANFVLMHYGTGAVMAVPAH-DQRDF-----EFAQKYSLP1KQVIAPLA 364
```

```
QY 104 DGNVKTAEVKEFGDTLVNRIELKGIIDFKEDGNILGHKLEYNYSNNVYIMADK-QKNGI 162
Db 365 DEEIDLTKQAFVEHGKLVNSDFDGKNF--DGAFFNG-----IADKLEKLGV 408
QY 163 ---KVNFKJRH-----NIEDSQVOLADHYQONTPIGDPVLLPDNHYL- 202
Db 409 GKRQVNYRLRDWGVSRQRYWGAPIMLTLLENGDVVFA-----PMEDLPIILPEDVMD 461
QY 203 STQSALSQKDPN 213
Db 462 GVKSPINADPN 472
RESULT 3
SYL_XYLFA
ID SYL_XYLFA STANDARD; PRT; 879 AA.
AC Q9PEG8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--trNA ligase) (LeuRS).
GN LEUS OR XF2176.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facinchi A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohne M., Fuzlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnsiegel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.B., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa."
RL Nature 406:151-159(2000).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF004031; AAF84975.1; ALT_INIT.
CC HAMAP; MF_00049; -; 1.
```





```

CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC   diphosphate + L-leucyl-tRNA(Leu).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to Class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE008725; AAL19599.1; -.
DR STyGene; SG77777; leus.
DR HAMAP; MF 00049; -. 1.
DR InterPro; IPR023202; Leu-tRNA-synt1a.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR009008; ValRS_1IERS_edit.
DR Pfam; PF001133; tRNA-synt_1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRfam; TIGR00396; leus_bact; 1.
DR PROSITE; PS00178; AA-tRNA LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
SQ SEQUENCE 860 AA; 96985 MW; D5003584DFBCCAB6 CRC64;

Query Match 6.9%; Score 87.5; DB 1; Length 860;
Best Local Similarity 23.3%; Pred. No. 14;
Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;

QY 50 TTGKLPVWPVTLVTTLSYGVCFSRYPDHMKQDFFKSAPEGVQERTIFFKDDGNYKT 109
D 314 TGEIIPV-WAANFVMEYGTGAVMAVEGH-DQRD-YEFASKYGLTIKPVILAADGSEPD 370
QY 110 RAEVKEGDTLVNRIELKGLDFKEDNGILGHKLEYNNHNVIMADKQKNGIKVNFKIR 169
D 371 SEQALTEKGVLFNSGEPDGLAFEAFAFNAIADKL-----AEKGVGERKVNRYLR 418
QY 170 H-----NIEDGSVOLADHYQONTPIGDGVPVLLPDNHYL-STQSALSKDP 212
D 419 DWGVSQRQYWGAPIPWVLTLEDGV-----LTPEDQLPVLIPDWDVMDGITSPIKADP 471

RESULT 6
ID_DPOE_YEAST STANDARD; PRT; 2222 AA.
AC P21951;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA polymerase epsilon, catalytic subunit A (EC 2.7.7.7) (DNA
DE polymerase II subunit A)
GN POL2 OR DUN2 OR YNL262W OR N0825.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1214-1221.
RX MEDLINE=90381771; PubMed=2169349;
RA Morrison A., Araki H., Clark A.B., Hamatake R.K., Sugino A.;
RT "A third essential DNA polymerase in S. cerevisiae.";
RL Cell 62:1143-1151(1990).
RN [2]
RP SEQUENCE OF 1-2221 FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=96310631; PubMed=8740425;
RA Sen-Gupta M., Lyck R., Fleig U., Niedenthal R.K., Hegemann J.H.;

```

```

RT "The sequence of a 24,152 bp segment from the left arm of chromosome
RT XIV from Saccharomyces cerevisiae between the BNI1 and the POL2
RT genes.";
RL Yeast 12:505-514 (1996).
RN [3]
RP TEMPERATURE SENSITIVE MUTANTS.
RX MEDLINE=92164663; PubMed=1537345;
RA Araki H., Ropp P.A., Johnson A.L., Johnston L.H., Morrison A.,
RA Sugino A.;
RT "DNA polymerase II, the probable homolog of mammalian DNA polymerase
RT epsilon, replicates chromosomal DNA in the yeast Saccharomyces
RT cerevisiae.";
RL EMBL J. 11:733-740 (1992).
CC -1- FUNCTION: DNA POLYMERASE II PARTICIPATES IN CHROMOSOMAL DNA
CC REPLICATION.
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA} (N).
CC -1- SUBUNIT: CONSISTS OF FIVE SUBUNITS (200 kDa, 80 kDa, 34 kDa, 30
CC kDa, AND 29 kDa).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: THE DNA POLYMERASE ACTIVITY DOMAIN RESIDES IN THE
CC N-TERMINAL HALF OF THE PROTEIN, WHILE THE C-TERMINUS IS NECESSARY
CC FOR COMPLEXING SUBUNITS B AND C.
CC -1- MISCELLANEOUS: In eukaryotes there are five DNA polymerases:
CC alpha, beta, gamma, delta, and epsilon which are responsible for
CC different reactions of DNA synthesis.
CC -1- SIMILARITY: Belongs to the DNA polymerase type-B family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M60416; AAA88711.1; -.
DR EMBL; X92494; CAA63235.1; -.
DR EMBL; Z71538; CAA96169.1; -.
DR PIR; A36028; A36028.
DR GeneOnline; 143268; -.
DR SGD; S0005206; POL2.
DR GO; GO:000731; P:DNA repair synthesis; IMP.
DR InterPro; IPR006172; DNA_pol_B.
DR InterPro; IPR006134; DNA_pol_B_dom.
DR InterPro; IPR006133; DNA_pol_B_exo.
DR Pfam; PF00136; DNA_pol_B; 1.
DR Pfam; PF03104; DNA_pol_B_exo; 1.
DR SMART; SM00486; POLBc; 1.
DR PROSITE; PS00116; DNA POLYMERASE B; FALSE NEG.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Zinc-finger; Nuclear protein.
FT ZN_FING 2108 2181 POTENTIAL.
FT VARIANT 644 644 M -> I (IN POL2-9 TS MUTANT).
FT VARIANT 710 710 P -> S (IN POL2-18 TS MUTANT).
SQ SEQUENCE 2222 AA; 255669 MW; CBCDD2AB147D65B CRC64;

Query Match 6.9%; Score 87.5; DB 1; Length 2222;
Best Local Similarity 28.2%; Pred. No. 44;
Matches 37; Conservative 14; Mismatches 49; Indels 31; Gaps 7;

QY 54 LPVPWP-TLVTTLSYGVCFSRYPDHM-----KQHDFFKSAPEGVV---QERTI 99
D 883 LPKSPFETFFTLNGLKLYSLVPCSLNRYVHQKFTNHQYQELKPLNVYIETHSENTI 942
QY 100 FFKDDGNYKTR--AEVFEFGDTLVNR-----IELKGDIFKEDGNILGHKLEYNN 147
D 943 FFEVDGPKAMILPSSKEGKIKKRAYAVFNEDGSLAELKGFELKRGEL---QLIKNFQ 999
QY 148 S--HNYVIMAD 156
D 1000 SDIFKVFLEGD 1010

```

```

RESULT 7
AC2L_HUMAN STANDARD; PRT; 689 AA.
ID QN0B1; Q81V99; Q8N234; Q96J11; Q96JX6; Q9NU28;
AC QN0B1; Q81V99; Q8N234; Q96J11; Q96JX6; Q9NU28;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Acetyl-coenzyme A synthetase 2-like, mitochondrial precursor
DE (BC 6.2.1.1) (Acetate--CoA ligase 2) (Acetyl-CoA synthetase 2).
GN ACAS2L OR KIAA1846.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagunley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Basley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harte J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Leharshaiho M.H., Leverhwa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConnachie L.J., McElay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkhen R., Sims S.,
RA Stuce C.D., Smith M.L., Soderlund C., Steward C.A., Suleston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RP [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Brain, and Testis;
RX MEDLINE=22388257; PubMed=12477932;
RA Krausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 169-689 FROM N.A. (ISOFORM 1).
RC TISSUE=Placenta, and Tongue;
RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,

```

```

RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamiyama K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Irie R., Sato H.,
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
RA Kikuchi H., Murakawa K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA Kawakami B., Ota T., Hayashi K., Sugiyama T., Otsuki T., Ishibashi T.,
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S.,
RA Kawai Y., Wakamatsu A., Kanehori K., Suzuki Y., Sugano S.,
RA Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 336-689 FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=21245130; PubMed=11347906;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XX.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.";
RL DNA Res. 8:85-95(2001).
CC -!- FUNCTION: Converts acetate to acetyl-CoA so that it can be used
for oxidation through the tricarboxylic cycle to produce ATP and
Co(2) (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + acetate + CoA = AMP + diphosphate +
acetyl-CoA.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Name=1;
CC IsoID=Q9NUB1-1; Sequence=Displayed;
CC Name=2;
CC IsoID=Q9NUB1-2; Sequence=VSP 007249;
CC Note=No experimental confirmation available;
CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
family.
CC -!- CAUTION: Ref.1 (CAB81884) sequence differs from that shown due to
erroneous gene model prediction.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to a
frameshift in position 250 and numerous sequencing errors.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL035661; CAB75500.1; -;
DR EMBL; AL080312; CAB81884.1; ALT_SEQ.
DR EMBL; BC039261; AAH39261.1; -;
DR EMBL; BC044588; AAH44588.1; -;
DR EMBL; AK027817; BAB55390.1; ALT_INIT.
DR EMBL; AK092295; BAC03853.1; ALT_SEQ.
DR EMBL; AB058749; BAB4475.1; -;
DR Genew; HGNC:16091; ACAS2L.
DR InterPro; IPR000873; AMP-bind.
DR Pfam; PF00501; AMP-binding; 1.
DR PROSITE; PS00455; AMP_BINDING; 1.
DR Ligase; Mitochondrial; Transit peptide; Alternative splicing.
KW TRANSIT 1 36 MITOCHONDRION (POTENTIAL)
FT CHAIN 37 689 ACETYL-COENZYME A SYNTHETASE 2-LIKE.
FT DOMAIN 45 53 POLY-ALA.
FT VARSPLOC 446 447 Missing (in isoform 2).
FT CONFLICT 277 277 /FTID=VSP 007249.
FT CONFLICT 488 488 V -> M (IN REF. 2; AAH39261).
FT CONFLICT 689 689 V -> M (IN REF. 2; AAH44588).
SQ SEQUENCE 689 AA; 74856 MW; 66E84E39302AD08B CRC64;
Query Match 6.8%; Score 87; DB 1; Length 689;
Best Local Similarity 24.1%; Pred. No. 12;
Matches 33; Conservative 16; Mismatches 36; Indels 52; Gaps 7;

```

```
QY 9 FTGVPILVELDGDVNGHKFSVSGEGGDATYQKLTALKFKITCTTGKLPVPMPTLVLTLSYG 68
DB 473 FFGIVPVLMDKGSV-----VEGNSVSGALCIS-----QAWPGWARTI--- 510
QY 69 VQCFSPYDPHMKQHDFFKSAMPQGYVQERIFPKDDGNKYTRA---EVKFEGETLVNRIE 125
DB 511 -----YGDHQRFVDAYFRAYP-GY-----YFTGDGAYRTEGGYQITGRMDVDVI----- 553
QY 126 LKGIKDFKEDGNILGHKL 142
DB 554 -----NISGHL 560

RESULT 8
SLAP_BACLI STANDARD; PRT; 874 AA.
AC P49052; 1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE S-layer protein precursor (Surface layer protein).
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NM 105;
RX MEDLINE=97082965; PubMed=8964497;
RA Zhu X., McVeigh R.R., Malathi P., Ghosh B.K.;
RT "The complete nucleotide sequence of the Bacillus licheniformis NM105
RT S-layer-encoding gene.";
RL Gene 173:189-194 (1996).
CC -!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly
CC of proteins which coat the surface of bacteria.
CC -!- SUBCELLULAR LOCATION: Cell wall.
CC -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U38842; AAC44405.1; -.
DR F01; JC4930; JC4930.
DR InterPro; IPR001119; SLH.
DR Pfam; PF00395; SLH; 3.
KW Signal; Cell wall; S-layer; Repeat.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 874 S-LAYER PROTEIN.
FT DOMAIN 33 93 SLH 1.
FT DOMAIN 94 155 SLH 2.
FT DOMAIN 156 217 SLH 3.
SQ SEQUENCE 874 AA; 92734 MW; EFADCA4FF2D7D32AF CRC64;
```

Query Match 6.8%; Score 87; DB 1; Length 874;  
Best Local Similarity 22.8%; Pred. No. 16;  
Matches 65; Conservative 28; Mismatches 96; Indels 96; Gaps 14;

```
QY 6 BELF--TGVVPII---VELDGVNGHKFSVSGEGDAT-----YCKLTLEFI 48
DB 551 KEVFPQTGVKVLVDVTTTNEGSGTSSIKVKGENVGAGTIHFQNPASGEGYGSLSHVEVT 610
QY 49 CTTGKLPVPMPTLV-----TTLSYGVQCFSPYDPHMKQHDFFKSAMPQGYVQER 97
DB 611 KSNIGHEAPRLVSVKAGQGEAADTLGNNTVAVQLNNTYTEGVYADADLAGY--RF 668
QY 98 TIFPKDDGNKYTRAEVKFEGETLVNRIELKGD---FKEDGNILGHKLYNNYNNHNYVM 154
DB 669 RV-----GNDKI-ASAKIEGKTLKVTGKTAGVTDVILTKDGATAGH-ATITVTQENIQIT 721
```

```
QY 155 ADKQK-----NGIKVNFKIRHN---IEDGSQLADH 182
DB 722 SVKEDVEVEQFENRKNVDRVLDVVKDKDVLNGIKLNISTEHKVRIVDEGTEQ----- 777
QY 183 YQONTPIGDPVLLPNDHYLSQTQSALSQDPNKRDRHMLVGLFVTA 227
DB 778 -----GKV-----YLDNRNDNATFDGND-----VALGYVTA 802
```

```
RESULT 9
CP51_CANGA STANDARD; PRT; 533 AA.
AC P50859; Q02312;
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 51 (EC 1.14.13.70) (CYPL1) (P450-LIA1) (Sterol 14-
DE alpha-demethylase) (Lanosterol 14-alpha demethylase) (P450-14DM).
GN ERG11 OR CYP51.
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2001-L5;
RX MEDLINE=96161286; PubMed=8593007;
RA Geber A., Hitchcock C.A., Swartz J.E., Pullen F.S., Marsden K.E.,
RA Kwon-Chung K.J., Bennett J.E.;
RT "Deletion of the Candida glabrata ERG3 and ERG11 genes: effect on cell
RT viability, cell growth, sterol composition, and antifungal
RT susceptibility.";
RL Antimicrob. Agents Chemother. 39:2708-2717 (1995).
RN [2]
RP SEQUENCE OF 60-473 FROM N.A.
RC STRAIN=ATCC 2001;
RX MEDLINE=95081364; PubMed=7989540;
RA Burgener-Kairuz P., Zuber J.P., Jaunin P., Buchman T.G., Bille J.,
RA Rossier M.;
RT "Rapid detection and identification of Candida albicans and
RT Torulopsis (Candida) glabrata in clinical specimens by
RT species-specific nested PCR amplification of a cytochrome P-450
RT lanosterol-alpha-demethylase (L1A1) gene fragment.";
RL J. Clin. Microbiol. 32:1902-1907 (1994).
CC -!- FUNCTION: Catalyzes C14-demethylation of lanosterol which is
CC critical for ergosterol biosynthesis. It transforms lanosterol
CC into 4,4'-dimethyl cholesta-8,14,24-triene-3-beta-ol (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: Obtusifoliol + 3 O(2) + 3 NADPH = 4-alpha-
CC methyl-5-alpha-ergosta-9,14,24(28)-trien-3-beta-ol + formate + 3
CC NADP(+) + 3 H(2)O.
CC -!- PATHWAY: Ergosterol biosynthesis.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L40389; AAB02329.1; -.
DR EMBL; S75389; AAB32679.1; -.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Electron transport; Oxidoreductase; Monooxygenase; Membrane; Heme;
KW Sterol biosynthesis; NADP.
FT METAL 472 472 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT CONFLICT 64 64 I -> M (IN REF. 2).
FT CONFLICT 473 473 I -> T (IN REF. 2).
```

```
SQ SEQUENCE 533 AA; 61305 MW; A0506C17507E6BF7 CRC64;
Query Match 6.8%; Score 86.5; DB 1; Length 533;
Best Local Similarity 21.4%; Pred. No. 9.5;
Matches 44; Conservative 32; Mismatches 81; Indels 49; Gaps 8;
QY 25 GHKFSVS--GEGEGDATYKGLTKFKICTTGTGLKLPVWPMTLVTLTSLYGVQCFSRYPDH--M 79
DB 109 GHEFFINAKLADVSAEAAASHL-----TTPVFGKVIYDCPNRLM 149
QY 80 KQHDFFKSA--PEGVY-----QRTIFPKDDGNKYKRAEVKFEQDGLVNRLELKGIDP 131
DB 150 EQKKFKVKGALTKFAFVRYPLIAEEIYKYFRNSKFKINENNISGIVDMVMSQPEM--TTF 207
QY 132 KEDGNILGHKLEYNYSNHYVIMADQKNGIKVNFKIRHNIEDGSVOLADHYQNTPIGD 191
DB 208 TARSLSLKGEMRDKLDTDFAYLSYDLDKGFTPIINF--VFNPLPLEHYRKRDHAQAQALS--- 263
QY 192 GPVLLPDNHYLSQSALSKDPNEKRD 217
DB 264 -----GTYSLSIKERREKND 278

RESULT 10
D152_HAEIN STANDARD; PRT; 795 AA.
AC P44935;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
GN H10917.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA McLagney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Spriggs T., Hedblom E., Cotton M.D.,
RA Fine L.D., Fritchman J.L., Nuyman D.T., Saudek D.M., Brandon R.C.,
RA Gnehm C.L., McDonald L.A., Fuhrmann J.L., Geoghagen N.S.M.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";
RL Science 269:496-512(1995).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: Belongs to the surface antigen D15 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; U32773; AAC22575.1; --
DR TIGR; H10917; --
DR InterPro; IPR000184; Bac_surfAg_D15.
DR Pfam; PF01103; Bac surface Ag; 1.
KW Antigen; Outer membrane; Signal; Complete proteome.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 795 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 795 AA; 87478 MW; B85691FC226ED44 CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 795;
Best Local Similarity 21.9%; Pred. No. 15;
Matches 48; Conservative 30; Mismatches 78; Indels 63; Gaps 11;
QY 65 LSYGVQCFSRYPDHMKQHDFF-----FKSAMPEGVYQE-----RTI 99
DB 427 IGYGTESGISYQASVKQDNFLGTGAASVITAGTKNDYGTSVNLGYTEPYFTTKDGVSLGGNV 486
QY 100 FFKDDGNKYKRAEVKFEQDGLVNRLELKGIDFKDEGNI---LCH-----KLEYNYS 148
DB 487 FFENDNSKSDTSSNKRRTYGSNTVL-GFPVNNNSYVVGCHYTNKISNFALEYN--- 542
QY 149 HNVYIMADQK-NGIKVNFKIRHNIEDGSVOLADHYQQ-----NTPIGDGFVLL 196
DB 543 RNLYIQSMKFKNGIKTN-----DPDFSGWYNLSNRGVFTPKGVKASLG-GRVTI 593
QY 197 P--DNHYLSQALSQKDPNEKRDHVVLLGFVTAAGITLG 233
DB 594 PGSDNKYKLSADVQGFYPLDRDLHVVWSAKASAGYANG 632

RESULT 11
D151_HAEIN STANDARD; PRT; 797 AA.
AC P46024;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype B;
RX MEDLINE=95255676; PubMed=7737523;
RA Flack F.S., Loosmore S., Chong P., Thomas W.R.;
RT "The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus influenzae."
RL Gene 156:97-99(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Minna / Serotype B, and Eagan / Serotype B;
RX MEDLINE=97427952; PubMed=9284140;
RA Loosmore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae species and may represent a universal protective antigen against invasive disease.";
RL Infect. Immun. 65:3701-3707(1997).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: Belongs to the surface antigen D15 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; U13961; AAA85645.1; --
DR EMBL; U0832; AAB61974.1; --
DR EMBL; U0833; AAB61976.1; --
DR PIR; JC4078; JC4078.
DR InterPro; IPR000184; Bac_surfAg_D15.
DR Pfam; PF01103; Bac surface Ag; 1.
KW Antigen; Outer membrane; Signal; POTENTIAL.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 797 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 797 AA; 87675 MW; 2F93DE53896AF1B CRC64;
```

```
DR InterPro; IPR009008; Valrs_1Iers_edit.
DR Pfam; PF00133; tRNA-synt 1; 1.
DR PRINTS; PRO0985; TRNASYNTHLEU.
DR TIGSFams; TIGR00396; leuS_bact; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_1; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 45 55 "HIGH" REGION.
FT SITE 637 641 "RMSKS" REGION.
FT BINDING 640 640 ATP (BY SIMILARITY).
SQ SEQUENCE 879 AA; 99823 MW; 4C2EE01B8FDC497E CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 879;
Best Local Similarity 22.2%; Pred. No. 17;
Matches 44; Conservative 28; Mismatches 69; Indels 57; Gaps 10;

QY 50 TTGKLPVPWPTLVTLISYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNY-- 107
DB 321 TNEQLPV-WVANFVLMAYGTGAVMAVPGHQDQDEF--ANKYGLPIRQVIALKEPKNQDE 377
QY 108 -----KTRAEVKPEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHNVI 153
DB 378 STWEPDVWRDWTYADKTR---EFE---LINSAEFDGLDYQGAPEVLAERFE----- 421
QY 154 MADKQKNG-IKVNFKIRHNIEDSGVOLADHYQONTPI-----GDGPFVLLPDN 199
DB 422 ---RQGRGQRRVNYRLR----DWGVSQRQYWGCPFVIVYCTGCGAVPVPENQLPVILPEN 474
QY 200 -HYLSTQSALSADPNEKR 216
DB 475 VAFSGTSGPIKTDPEWRK 492

RESULT 13
VIT4_CAEL
ID VIT4_CAEL STANDARD; PRT; 1603 AA.
AC P18947; Q9BPP3;
DT 01-NOV-1990 (Rel. 16, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vitellogenin 4 precursor.
GN VIT-4 OR F59D8.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE OF 1-282 FROM N.A.
RA Blumenthal T., Spieth J., Zucker E.;
RL Submitted (NOV-1986) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE OF 1-71 FROM N.A.
RX MEDLINE=85269643; PubMed=4022780;
RA Spieth J., Denison K., Kirtland S., Cane J., Blumenthal T.;
RT "The C. elegans vitellogenin genes: short sequence repeats in the
RT promoter regions and homology to the vertebrate genes.";
RL Nucleic Acids Res. 13:5283-5295(1985).
CC -!- FUNCTION: Precursor of the egg-yolk proteins that are sources of
CC nutrients during embryonic development (Potential).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Synthesized in Caenorhabditis only by 32 cells
CC building the intestine of adult hermaphroditic individuals; they
CC are coranationally secreted into the body cavity and
CC subsequently taken up by the gonad.
CC -!- SIMILARITY: Contains 1 WVF domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AC024137; AAK09074.1; --  
DR EMBL; M11498; AAA28163.1; --  
DR EMBL; X02754; CAA36531.1; --  
DR FIK; A43084; A43084.  
DR WormPep; F59D8.2; CE26817.  
DR InterPro; IPR001747; Lipid\_transprt\_N.  
DR InterPro; IPR001846; VWF\_D.  
DR Pfam; PF01347; Vitellogenin\_N; 1.  
DR SMART; SM00638; LPD\_N; 1.  
DR SMART; SM00216; VMD; 1.  
KW Storage protein; Multigene family; Signal.  
FT SIGNAL 1 15 POTENTIAL.  
FT CHAIN 16 1603 VITELLOGENIN 4.  
FT DOMAIN 1308 1455 VWF\_D.  
FT CONFLICT 30 30 Y -> V (IN REF. 3).  
FT CONFLICT 169 169 L -> V (IN REF. 2).  
FT CONFLICT 183 187 EVAYT -> RSRUH (IN REF. 2).  
FT CONFLICT 275 275 T -> S (IN REF. 2).  
SQ SEQUENCE 1603 AA; 186307 MW; E303170325BC99BB CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 1603;  
Best Local Similarity 23.4%; Pred. No. 36;  
Matches 52; Conservative 32; Mismatches 69; Indels 69; Gaps 12;  
QY 1 MVSKGSELFVGVPIILVELDGDVNGHKFSVSGEGGDATYGLTKLTKPCTTGKLPVWPPT 60  
DB 162 MESDKDSLFLNVHEKTMGDCV-AYTIVQSG-GKTIYKSVNFDKCIIR-----PE 211  
QY 61 LVTTLSYGVQCFSRYPDHMKQHDFKSAPEG-YVQERTIF----FKDDG-----105  
DB 212 TAYGLAFGSEC-----KECEKEGVQVQPTVYTFYTFKNEKLQSEVNSIYT 257  
QY 106 -----NYKTRAEVKRPGDGLVNRILKIGDFKEDGNILGHKLEYNVNSHNVIMAD 156  
DB 258 LNVNGQEVVKSRAKTVFVEESKINR-EIK-----KVGPKKEIIVSMENKLEIQ 308  
QY 157 KQKNG-----IKVNFKIRHNIEDGVSQVLADHYQQNTP 188  
DB 309 FYKQGDKAENVPPKAIIEQKV-EQLKEIFRQIQEH-EQNTP 348

RESULT 14  
TRNA CAMJE  
ID TRNA CAMJE STANDARD; PRT; 357 AA.  
AC Q9PP92;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE tRNA (Uracil-5)-methyltransferase (EC 2.1.1.35) (tRNA(M-5-U54)-  
DE methyltransferase) (RUMT).  
GN TRNA OR CJO831C.  
OS Campylobacter jejuni.  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
OC Campylobacteraceae; Campylobacter.  
OX NCBI\_TaxID=197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCTC 11168;  
RX MEDLINE=20150912; PubMed=10688204;  
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,  
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
RA Whitehead S., Barrall B.G.;  
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
RT reveals hypervariable sequences."  
RL Nature 403:665-668(2000).

CC -!- FUNCTION: Catalyzes the formation of 5-methyl-uridine at position  
CC 54 (M-5-U54) in all tRNA (By similarity).  
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-  
CC homocysteine + tRNA containing thymine.  
CC -!- SIMILARITY: Belongs to the RNA MSU methyltransferase family. Trna  
CC subfamily.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AL139076; CAB73096.1; --  
DR FIK; G81355; G81355.  
DR HAMAP; MF\_01011; --; 1.  
DR InterPro; IPR000051; SAM bind.  
DR InterPro; IPR001566; TrmA.  
DR PROSITE; PS01230; TRMA\_1; 1.  
DR PROSITE; PS01231; TRMA\_2; FALSE NEG.  
KW Transferase; Methyltransferase; tRNA processing; Complete proteome.  
FT DOMAIN 207 213 S-ADENOSYLMETHIONINE BINDING (BY  
FT ACT SITE 315 315 BY SIMILARITY.  
FT SEQUENCE 357 AA; 42276 MW; CEC5328347CEB497 CRC64;

Query Match 6.8%; Score 86; DB 1; Length 357;  
Best Local Similarity 24.8%; Pred. No. 6.5;  
Matches 30; Conservative 18; Mismatches 39; Indels 34; Gaps 5;  
QY 80 KOHDFPKSAMPGYVOERTIFFPKDGNKYKTRAEVKF--EGDTLV-----NRIELKG 128  
DB 14 EKHSFKYKFKFYKDFKPKASKDKHYTRAEISFYHENDTLFYAMFDPKSKKYIIEY 73  
QY 129 IDPKED-----GNILGHKLEYNVNSHNVIMADKQKNGIKVNFKIRHNIE 173  
DB 74 LDFADEKICAFMPRLLEYLRQDNKLEKL-----FGVEFLTQKQ--LSITLLYHKNIE 125  
QY 174 D 174  
DB 126 D 126

RESULT 15  
DPOL\_HPBHE  
ID DPOL\_HPBHE STANDARD; PRT; 788 AA.  
AC F13846;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE P protein [Includes: DNA-directed DNA polymerase (EC 2.7.7.7); RNA-  
DE directed DNA polymerase (EC 2.7.7.49); Ribonuclease H (EC 3.1.26.4)].  
GN P.  
OS Heron hepatitis B virus.  
OC Viruses; Retrovird viruses; Hepadnaviridae; Avihepadnavirus.  
OX NCBI\_TaxID=28300;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88333160; PubMed=3418788;  
RA Sprengel R., Kaleta E.F., Will H.;  
RT "Isolation and characterization of a hepatitis B virus endemic in  
RT herons."  
RL J. Virol. 62:3832-3839(1988).  
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
CC + {DNA} [N].  
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-  
CC phosphomonoester.  
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR ENBL; M22056; AAA45738.1; -;  
DR PIR; A30082; JDLVHH.  
DR InterPro; IPR001462; DNAPol\_viral\_C.  
DR InterPro; IPR000201; DNAPol\_viral\_N.  
DR InterPro; IPR000477; RVTse.  
DR Pfam; PF00336; DNA\_pol\_viral\_C; 1.  
DR Pfam; PF00242; DNA\_pol\_viral\_N; 1.  
DR Pfam; PF00078; rvt; 1.  
DR ProDom; PD000814; DNAPol\_viral\_C; 1.  
DR Transferrase; RNA-directed DNA polymerase; DNA-directed DNA polymerase;  
KW Hydrolase; Nuclease; Endonuclease; DNA replication; DNA-binding.  
SQ SEQUENCE 788 AA; 90070 MW; FB44F38F75EADF44 CRC64;

Query Match 6.7%; Score 85.5; DB 1; Length 788;  
Best Local Similarity 19.7%; Pred. No. 18;  
Matches 40; Conservative 32; Mismatches 66; Indels 65; Gaps 10;  
QY 58 WPTLVTTLSYGVCFSRYPDHMKQH-----DPEKSAMPEGYVOERT-----IPFKDDGNYK 108  
DB 139 WPKSISYLPVHSGVKPKYPEFQNHESLVNDYLNKLFEGILYKRVSKHLVTFK--GPYF 196  
QY 109 T-----RAEVKFGDTLVNRIELKGIDFKEDGNILGHKLEYNYNHNVYIM 154  
DB 197 TWEQKHLVPOQHGYASKINDROESRRRIITATSSRKNDSSRI-----FGAHN----- 245  
QY 155 ADKQKNGIKVNFKIRHNIEDGVSQVLADHYQ-----QNTPIGDGPFLL-PDNHYL 202  
DB 246 -----NGRKISY---HSTRDGSRLSGRTSDPTSRGALAGDSDTPIGPGSTAAHPSTHHV 297  
QY 203 STQ-----SALSKDPNEKR 216  
DB 298 DRRRQKGGVLOAISRPESETR 320

Search completed: June 21, 2004, 15:55:22  
Job time : 7.55556 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:45:30 ; Search time 30.6667 Seconds  
(without alignments)  
2458.984 Million cell updates/sec

Title: US-09-887-784-222L  
Perfect score: 1274  
Sequence: 1 MVSKEELFTGVVILVELD.....VLLGFVTAAGITLGMDELYK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues 1017041

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL\_25.\*
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phase.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_virus.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_unclassified.\*
  - 15: sp\_rvirus.\*
  - 16: sp\_bacteriap.\*
  - 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1243	97.6	238	2 Q8GHE2	Q8ghe2 azotobacter
2	1240	97.3	238	5 Q93125	Q93125 aequorea vi
3	1238	97.2	238	2 Q8GHE4	Q8ghe4 azomonas ag
4	1237	97.1	238	2 Q8GHE3	Q8ghe3 azotobacter
5	1205	94.6	238	5 Q17105	Q17105 aequorea vi
6	1190	93.4	238	5 Q17106	Q17106 aequorea vi
7	1085	85.2	238	5 Q8WTC6	Q8wtc6 aequorea ma
8	1081	84.9	238	5 Q8WPC95	Q8wpc95 aequorea ma
9	1077	84.5	238	5 Q8WTC4	Q8wtc4 aequorea ma
10	1075	84.4	238	5 Q8WTD0	Q8wtd0 aequorea ma
11	1074	84.3	238	5 Q8WTC8	Q8wtc8 aequorea ma
12	1074	84.3	238	5 Q8WTC9	Q8wtc9 aequorea ma
13	1072	84.1	238	5 Q8WTC7	Q8wtc7 aequorea ma
14	1070	84.0	238	5 Q8WTC5	Q8wtc5 aequorea ma
15	252.5	19.8	225	5 Q95UA7	Q95ua7 montastraea
16	252.5	19.8	225	5 Q7Z0W5	Q7z0w5 montastraea

17	247	19.4	225	5	Q963F5	Q963f5 montastraea
18	244.5	19.2	236	5	Q8T6U0	Q8t6u0 dendronepht
19	242.5	19.0	225	5	Q7Z0W9	Q7z0w9 montastraea
20	240	18.8	225	5	Q8I6J8	Q8i6j8 trachyphyl
21	238.5	18.7	266	5	Q9U6Y3	Q9u6y3 clavularia
22	233	18.3	225	5	Q7Z0W4	Q7z0w4 montastraea
23	232	18.2	224	5	Q8MU48	Q8mu48 montastraea
24	232	18.2	225	5	Q8T5F1	Q8t5f1 montastraea
25	214	16.8	227	5	Q7Z0W6	Q7z0w6 montastraea
26	214	16.8	234	5	Q7Z0W7	Q7z0w7 montastraea
27	212.5	16.7	259	5	Q8NMA2	Q8nma2 agaricia fr
28	212	16.6	239	5	Q8NMA1	Q8nma1 agaricia ag
29	209	16.4	234	5	Q8T5F2	Q8t5f2 montastraea
30	209	16.4	234	5	Q8MU47	Q8mu47 montastraea
31	208.5	16.4	229	5	Q9U6Y6	Q9u6y6 anemonia ma
32	206	16.2	227	5	Q962P9	Q962p9 montastraea
33	206	16.2	227	5	Q7Z0W8	Q7z0w8 montastraea
34	205.5	16.1	232	5	Q9GP15	Q9gpi5 anemonia su
35	204.5	16.1	214	5	Q86LV7	Q86lv7 meandrina m
36	204	16.0	221	5	Q95P04	Q95p04 goniopora t
37	203.5	16.0	214	5	Q86LV8	Q86lv8 meandrina m
38	203.5	16.0	238	5	Q9BLV9	Q9bly9 renilla mue
39	202	15.9	227	5	Q95VT0	Q95vt0 montastraea
40	201.5	15.8	225	5	Q9U6Y8	Q9u6y8 discosoma s
41	201.5	15.8	232	5	Q9GZ28	Q9gz28 anemonia su
42	198.5	15.6	222	5	Q7Z168	Q7z168 cerianthus
43	198.5	15.6	225	5	Q8T6T9	Q8t6t9 radianthus
44	198	15.5	235	5	Q8T5F0	Q8t5f0 scolymia cu
45	197.5	15.5	232	5	Q9U6Y7	Q9u6y7 discosoma s

ALIGNMENTS

RESULT 1

Q8GHE2 ID Q8GHE2 PRELIMINARY; PRT; 238 AA.  
AC Q8GHE2;  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Green fluorescence protein.  
GN 2289GFP.  
OS Azotobacter vinelandii.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Azotobacter.  
OX NCBI\_TaxID=354;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM2289;  
RA Koranyi P., Berenyi M., Burg K.;  
RT "Occurrence of green fluorescence protein in diazotrophic bacteria  
RL Azomonas and Azotobacter.";  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF324408; AAN86140.1; -;  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR009017; GFP like.  
DR InterPro; IPR000786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFLUORSCENT.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
SQ SEQUENCE 238 AA; 26914 MW; F84840F1F9064018 CRC64;

Query Match 97.6%; Score 1243; DB 2; Length 238;  
Best Local Similarity 97.9%; Pred. No. 1.9e-96;  
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 VSKGELFTGVVILVELDGVNGHKFSVSGEGDATYKGLTKLKEICTTGKLPVWPPTL 61  
Db 1 MKGKELFTGVVILVELDGVNGHKFSVSGEGDATYKGLTKLKEICTTGKLPVWPPTL 60  
Qy 62 VTTLSYGVQCFRYPDPMKQHDFFKSAMPEGYVQERTIFFKDDGNVYKTRAEVKEGDTLLV 121

```

Db      61 VTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
QY      122 NRIELKGIDFKEDGNILGHKLEYNYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLAD 181
Db      121 NRIELKGIDFKEDGNILGHKLEYNYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLAD 180
QY      182 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVTTAAGITLGMDELYK 239
Db      181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVTTAAGITLGMDELYK 238

RESULT 2
Q93125 ID Q93125 PRELIMINARY; PRT; 238 AA.
AC Q93125
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein mutant 3.
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Chordata; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96305137; PubMed=8707053;
RA Cormack B.P., Valdivia R.H., Falkow S.;
RT "FACS-optimized mutants of the green fluorescent protein (GFP).";
RL Gene 173:33-38(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Cormack B.P., Bertram G., Egerton M., Cow N.A.R., Falkow S.,
RA Brown A.J.P.;
RT "Yeast Enhanced Green Fluorescent Protein (yEGFP): a reporter of gene
RT expression in Candida albicans.";
RL Microbiology 0:0-0(1996).
DR EMBL; U73901; AAB18957.1; -.
DR HSP; P42212; IBBP.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR Pfam; PF01353; GFP.1.
DR PRINTS; PR01229; GFP.1.
DR PRODOM; PD013756; Green_fl_protein.1.
SQ SEQUENCE 238 AA; 28640 MW; A28622809A9DEA60 CRC64;

Query Match 97.3%; Score 1240; DB 5; Length 238;
Best Local Similarity 97.3%; Pred. No. 3.4e-96;
Matches 232; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSKGELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKLFCITCTGKLPVWPPTL 61
:
1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKGLTKLFCITCTGKLPVWPPTL 60

Db      62 VTTLSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
Db      61 VTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
QY      122 NRIELKGIDFKEDGNILGHKLEYNYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLAD 181
Db      121 NRIELKGIDFKEDGNILGHKLEYNYNHNHYIMADKQNGIKVNFKIRHNIEDGSVQLAD 180
QY      182 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVTTAAGITLGMDELYK 239
Db      181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVTTAAGITLGMDELYK 238

RESULT 4
Q8GHE3 ID Q8GHE3 PRELIMINARY; PRT; 238 AA.
AC Q8GHE3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescence protein.
GN 85GFP.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
RT Azomonas and Azotobacter.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324406; AAN86138.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP.1.
DR PRINTS; PR01229; GFP.1.
DR PRODOM; PD013756; Green_fl_protein.1.
SQ SEQUENCE 238 AA; 26902 MW; 15FE9B9C5B4F6B89 CRC64;

Query Match 97.1%; Score 1237; DB 2; Length 238;

```

RESULT 6  
Q17106  
ID Q17

```
DR PRINTS; PR01229; GFP, FLUORESCENT.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27015 MW; 658FD75E8926903 CRC64;

Query Match      85.2%; Score 1085; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 3.6e-83;
Matches 196; Conservative 20; Mismatches 22; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDGYKLTGKLFICITGKLPVWPPTL 61
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSKGEELFTGIVPVLIELDGDVGHKFSVRGEGDADYKLEIKFICTTGKLPVWPPTL 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 62 VTTLSGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFPKDDGNKYKTRAEVKEGDTLV 121
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VTTLSYGIQCFARYPEHMKWDFKSAPEGYIQERTIFFQDDGKYKTRGEVKEGDTLV 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 122 NRIELKGIDFKEDGNILGHKLEYNFNHNYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 NRIELKGMDPKEDGNILGHKLEYNFNHNYIMPDKANNGLKYNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 HYQTNVPLGDGPVLIPINHLYSTQTAISKDRNETRDHMLVFLFFSACGHTGMDLYK 238

RESULT 8
Q8WP95 PRELIMINARY; PRT; 238 AA.
AC Q8WP95;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFPXM.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qian Y.X., Pang S.Q.,
  Li S.J., Xia N.S.;
  Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY013824; AAK02062.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP, 1.
DR PRINTS; PR01229; GFP, 1.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27049 MW; 8185D0E5E29012B CRC64;

Query Match      84.9%; Score 1081; DB 5; Length 238;
Best Local Similarity 81.9%; Pred. No. 7.7e-83;
Matches 195; Conservative 20; Mismatches 23; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDGYKLTGKLFICITGKLPVWPPTL 61
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSKGEELFTGIVPVLIELDGDVGHKFSVRGEGDADYKLEIKFICTTGKLPVWPPTL 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 62 VTTLSGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFPKDDGNKYKTRAEVKEGDTLV 121
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VTTFSYGIQCFARYPEHMKWDFKSAPEGYIQERTIFFQDDGKYKTRGEVKEGDTLV 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 122 NRIELKGIDFKEDGNILGHKLEYNFNHNYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 NRIELKGMDPKEDGNILGHKLEYNFNHNYIMPDKANNGLKYNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 HYQTNVPLGDGPVLIPINHLYSTQTAISKDRNETRDHMLVFLFFSACGHTGMDLYK 238
```

```
RESULT 9
Q8WTC4 PRELIMINARY; PRT; 238 AA.
AC Q8WTC4;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Li S.J., Xia N.S.;
  Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL33918.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP, 1.
DR PRINTS; PR01229; GFP, 1.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match      84.5%; Score 1077; DB 5; Length 238;
Best Local Similarity 82.8%; Pred. No. 1.7e-82;
Matches 197; Conservative 17; Mismatches 24; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDGYKLTGKLFICITGKLPVWPPTL 61
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSKGEELFTGVVPIVLVDGDNVGHKFSVRGEGDADYKLEIKFICTTGKLPVWPPTL 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 62 VTTLSGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFPKDDGNKYKTRAEVKEGDTLV 121
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VTTLSYGIQCFARYPEHMKWDFKSAPEGYIQERTIFFQDDGKYKTRGEVKEGDTLV 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 122 NRIELKGIDFKEDGNILGHKLEYNFNHNYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 NRIELKGMDPKEDGNILGHKLEYNFNHNYIMPDKANNGLKYNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 HYQTNVPLGDGPVLIPINHLYSTQTAISKDRNETRDHMLVFLFFSACGHTGMDLYK 238

RESULT 10
Q8WTD0 PRELIMINARY; PRT; 238 AA.
AC Q8WTD0;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
  Li S.J., Xia N.S.;
  Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL33918.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP, 1.
DR PRINTS; PR01229; GFP, 1.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match      84.5%; Score 1077; DB 5; Length 238;
Best Local Similarity 82.8%; Pred. No. 1.7e-82;
Matches 197; Conservative 17; Mismatches 24; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDGYKLTGKLFICITGKLPVWPPTL 61
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSKGEELFTGVVPIVLVDGDNVGHKFSVRGEGDADYKLEIKFICTTGKLPVWPPTL 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 62 VTTLSGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFPKDDGNKYKTRAEVKEGDTLV 121
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VTTLSYGIQCFARYPEHMKWDFKSAPEGYIQERTIFFQDDGKYKTRGEVKEGDTLV 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 122 NRIELKGIDFKEDGNILGHKLEYNFNHNYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 NRIELKGMDPKEDGNILGHKLEYNFNHNYIMPDKANNGLKYNFKIRHNIEGGVOLAD 180

QY 182 HYQONTPIGDGPVLLPDNHYLSQSALSKDPNEKRDHMLVLLGFVTAAGITLGMDELYK 239
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 HYQTNVPLGDGPVLIPINHLYSTQTAISKDRNETRDHMLVFLFFSACGHTGMDLYK 238
```

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF435427; AAL33912.1; -.  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR009017; GFP like.  
DR InterPro; IPR000786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFLUORESCENT.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
SQ SEQUENCE 238 AA; 26997 MW; 5F80A192173CB84D CRC64;

Query Match 84.4%; Score 1075; DB 5; Length 238;  
Best Local Similarity 81.5%; Pred. No. 2.5e-82;  
Matches 194; Conservative 20; Mismatches 24; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVWPILVELDGVNGHKFSVSGEGDATYKGLTKLFICITTKGLPVPWPTL 61  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
1 MSKGEELFTGIVPVLIELDGDVGHKFSVRGEGDADYKLEIKFICITTKGLPVPWPTL 60  
QY 62 VTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLV 121  
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
61 VTTLGVIQCFARYPEHMKMNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGLTLV 120  
QY 122 NRLELKGIDFKEDGNILGHKLEYNYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181  
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
121 NRLELKGMDFKEDGNILGHKLEYNFNHNYIMPDKANGLKYNFKIRHNIEGGVOLAD 180  
QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239  
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
181 HYQTNVPLGDGPVLLPINHYLSQTALSKDRNETRDMHVFLEFFSACGHTGMDLYK 238

## RESULT 11

Q8WTC8 PRELIMINARY; PRT; 238 AA.

AC Q8WTC8; (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Green fluorescent protein.  
GN Aequorea macrodactyla.  
OS Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;  
OC Aequoreidae; Aequorea.  
OX NCBI\_TaxID=147615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GFPxm163;  
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,  
RA Li S.J., Xia N.S.;  
RT "Colorful mutants of green fluorescent protein from Aequorea  
RT macrodactyla.";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF435429; AAL33914.1; -.  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR009017; GFP like.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFLUORESCENT.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;

Query Match 84.3%; Score 1074; DB 5; Length 238;

Best Local Similarity 81.5%; Pred. No. 3e-82;  
Matches 194; Conservative 20; Mismatches 24; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVWPILVELDGVNGHKFSVSGEGDATYKGLTKLFICITTKGLPVPWPTL 61  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
1 MSKGEELFTGIVPVLIELDGDVGHKFSVRGEGDADYKLEIKFICITTKGLPVPWPTL 60  
QY 62 VTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLV 121  
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
61 VTTLGVIQCFARYPEHMKMNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGLTLV 120

QY 122 NRLELKGIDFKEDGNILGHKLEYNYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181  
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
121 NRLELKGMDFKEDGNILGHKLEYNFNHNYIMPDKANGLKYNFKIRHNIEGGVOLAD 180  
QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239  
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
181 HYQTNVPLGDGPVLLPINHYLSQTALSKDRNETRDMHVFLEFFSACGHTGMDLYK 238

## RESULT 12

Q8WTC9 PRELIMINARY; PRT; 238 AA.

AC Q8WTC9; (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Green fluorescent protein.  
GN Aequorea macrodactyla.  
OS Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;  
OC Aequoreidae; Aequorea.  
OX NCBI\_TaxID=147615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GFPxm162;  
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,  
RA Li S.J., Xia N.S.;  
RT "Colorful mutants of green fluorescent protein from Aequorea  
RT macrodactyla.";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF435428; AAL33913.1; -.  
DR GO; GO:0006091; P:energy pathways; IEA.  
DR InterPro; IPR009017; GFP like.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
SQ SEQUENCE 238 AA; 27031 MW; 5F80A19C19DC584D CRC64;

Query Match 84.3%; Score 1074; DB 5; Length 238;

Best Local Similarity 81.5%; Pred. No. 3e-82;  
Matches 194; Conservative 20; Mismatches 24; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVWPILVELDGVNGHKFSVSGEGDATYKGLTKLFICITTKGLPVPWPTL 61  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
1 MSKGEELFTGIVPVLIELDGDVGHKFSVRGEGDADYKLEIKFICITTKGLPVPWPTL 60  
QY 62 VTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLV 121  
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
61 VTTLGVIQCFARYPEHMKMNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGLTLV 120  
QY 122 NRLELKGIDFKEDGNILGHKLEYNYNHNHNYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181  
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
121 NRLELKGMDFKEDGNILGHKLEYNFNHNYIMPDKANGLKYNFKIRHNIEGGVOLAD 180  
QY 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLGFTVTAAGITLGMDELYK 239  
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
181 HYQTNVPLGDGPVLLPINHYLSQTALSKDRNETRDMHVFLEFFSACGHTGMDLYK 238

## RESULT 13

Q8WTC7 PRELIMINARY; PRT; 238 AA.

AC Q8WTC7; (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Green fluorescent protein.  
GN Aequorea macrodactyla.  
OS Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;  
OC Aequoreidae; Aequorea.  
OX NCBI\_TaxID=147615;



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:41:49 ; Search time 47.1111 Seconds  
(without alignments)  
1433.395 Million cell updates/sec

Title: US-09-887-784-222V

Perfect score: 1274

Sequence: 1 MVSKEELFTGVVILVELD.....VLVGFVTAAGITGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: Genesecp1980s:\*
- 2: Genesecp1990s:\*
- 3: Genesecp2000s:\*
- 4: Genesecp2001s:\*
- 5: Genesecp2002s:\*
- 6: Genesecp2003as:\*
- 7: Genesecp2003bs:\*
- 8: Genesecp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1271	99.8	239	5 AAE17518	Aae17518 Enhanced
2	1271	99.8	363	6 ABR40352	Abr40352 Human ami
3	1271	99.8	893	4 AAG65781	Aag65781 Amino aci
4	1271	99.8	1132	4 AAG65782	Aag65782 Amino aci
5	1263	99.1	239	5 AAE17517	Aae17517 Enhanced
6	1260	98.9	239	3 AAB22882	Aab22882 Enhanced
7	1260	98.9	239	3 AAY54349	Aay54349 Amino aci
8	1260	98.9	239	3 AAY79584	Aay79584 EGFP sign
9	1260	98.9	239	4 AAE50804	Aab50804 Jellyfish
10	1260	98.9	239	4 AAB85900	Aab85900 A. victor
11	1260	98.9	239	4 AAB31171	Aab31171 Amino aci
12	1260	98.9	239	5 AAG66198	Aag66198 A. victor
13	1260	98.9	239	5 AAG94444	Abg94444 Protease
14	1260	98.9	239	5 AAE14599	Aae14599 Aequorea
15	1260	98.9	239	6 AAE34958	Aae34958 Aequorea
16	1260	98.9	239	6 AAG79829	Aag79829 Green flu
17	1260	98.9	239	6 ABR83616	Abr83616 Green flu
18	1260	98.9	239	6 ADA38074	Ada38074 Aequorea
19	1260	98.9	239	7 ABU63204	Abu63204 Aequorea
20	1260	98.9	239	7 ADC18358	Adc18358 EGFP (enh
21	1260	98.9	239	7 ABW00914	Abw00914 Aequorea
22	1260	98.9	239	7 ADE28570	Ade28570 Enhanced
23	1260	98.9	246	7 ABM79011	Abm79011 Enhanced
24	1260	98.9	248	5 AAG68319	Aag68319 Jellyfish
25	1260	98.9	259	5 AAU99804	Aau99804 Biomembra

ALIGNMENTS

RESULT 1

AAE17518  
ID AAE17518 standard; protein; 239 AA.

XX  
AC AAE17518;

DT XX  
XX 22-APR-2002 (first entry)

DE Enhanced F64L-E222G jellyfish green fluorescent protein mutant.

XX Jellyfish; green fluorescent protein; GFP; protein redistribution;  
KW cellular function; genetic reporter; mutant; Stoke's shift; muten.

XX Aequorea victoria.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 65

FT /note= "Wild type Phe substituted with Leu; This  
corresponds to position 64 in the wild type protein"

FT Misc-difference 223

FT /note= "Wild type Glu substituted with Gly; This  
corresponds to position 222 in the wild type protein"

FT WO200198338-A2.

XX 27-DEC-2001.

XX 18-JUN-2001; 2001WO-BP006848.

XX 19-JUN-2000; 2000DK-00000953.

PR 20-JUN-2000; 2000US-0212681P.

PR 10-MAY-2001; 2001DK-00000739.

PR 10-MAY-2001; 2001US-0290170P.

XX (BIOI-) BIOIMAGE AS.

XX Bjorn SP, Pagliaro L, Thastrup O;

XX WPI; 2002-098224/13.

DR N-PSDB; AAD28163.

XX Novel fluorescent protein in in vitro assay for measuring protein kinase  
PT activity or dephosphorylation activity, or for measuring protein  
PT redistribution, has a green fluorescent protein with F64L and E222G  
PT mutation.

XX Claim 9; Page 37; 41pp; English.

XX The invention relates to a fluorescent protein derived from green  
CC fluorescent protein (GFP) or its analogue. The GFP containing mutations  
CC at F64L and E222G has a bigger compared to other GFP's making it very  
CC suitable for high throughput screening due to better resolution. The  
CC fluorescent protein is useful in vitro assays for measuring protein  
CC kinase activity or dephosphorylation activity, or for measuring protein  
CC redistribution. The fluorescent protein is useful in studying cellular  
CC functions in living cells; as protein tags in transgenic animals, living  
CC and fixed cells; organelle tags, secretion marker and genetic reporter.  
CC The fluorescent protein is also useful as a cell or organelle integrity  
CC marker, a marker for changes in cell morphology, as transfection marker,  
CC and as a marker to be used in combination with fluorescence activated  
CC cell sorting (FACS). The novel proteins can also be used as reporters to  
CC monitor live or dead biomass of organisms, such as fungi. The fluorescent  
CC protein is also useful as markers in transcriptional and translational  
CC fusions for performing transposon vector mutagenesis and as a reporter  
CC for bacterial detection. Transposons encoding the fluorescent protein are  
CC useful for screening promoters and for tagging plasmids and chromosomes.  
CC The fluorescent protein engineered into the genome of a phage is useful  
CC for designing diagnostic tool. The present sequence is a DNA encoding  
CC enhanced F64L-E222G jellyfish green fluorescent protein (GFP) mutant  
XX  
SQ Sequence 239 AA;

Query Match 99.8%; Score 1271; DB 5; Length 239;  
Best Local Similarity 99.8%; Pred. No. 1.1e-122;  
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSKEBELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60  
DB 1 MVSKEBELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60  
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQNTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239  
DB 181 DHYQNTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239

RESULT 2  
ABR40352  
ID ABR40352 standard; protein; 363 AA.  
XX  
AC ABR40352;  
XX  
DT 08-JUL-2003 (first entry)  
XX  
DE Human amino acid sequence SEQ ID NO: 6.  
XX  
DE Human; heterologous conjugate; intracellular protein.  
XX  
OS Homo sapiens.  
OS Aequoria victoria.  
XX  
PN WO2003029827-A2.  
XX  
PD 10-APR-2003.  
XX  
PF 01-OCT-2002; 2002WO-DK000651.  
XX  
PR 01-OCT-2001; 2001DK-0001433.  
PR 11-OCT-2001; 2001US-032896P.  
XX  
PA (BIOL-) BIOWAGE AS.  
XX  
PI Terry BR, Nielsen SJ;

XX WPI: 2003-430211/40.  
DR N-PSDB; ACC72604.  
XX  
PT Novel cell for identifying modulators of protein interaction, contains a  
PT first conjugate comprising anchor protein, second conjugate having type B  
PT interactor protein and third conjugate with detectable group.  
XX  
PS Disclosure; Page 112-113; 118pp; English.  
XX  
CC The invention relates to a novel cell, comprising three heterologous  
CC conjugates (HC), a first HC (HC1) comprising an anchor protein that  
CC specifically binds to an internal structure within the cell conjugated to  
CC an interactor protein (IP) of type A, a second HC (HC2) comprising IP of  
CC type B conjugated to a first protein of interest, and a third HC (HC3)  
CC comprising a second protein of interest conjugated to detectable group.  
CC The cell is useful for detecting if a compound disrupts or induces the  
CC interaction between two intracellular proteins. The cell is also useful  
CC for screening compounds that modulate the interaction between two  
CC intracellular proteins. The present sequence is used in the  
CC exemplification of the invention  
XX  
SQ Sequence 363 AA;

Query Match 99.8%; Score 1271; DB 6; Length 363;  
Best Local Similarity 99.6%; Pred. No. 2e-122;  
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSKEBELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60  
DB 1 MVSKEBELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPT 60  
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQNTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239  
DB 181 DHYQNTPIGDGVPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239

RESULT 3  
AAG65781  
ID AAG65781 standard; protein; 893 AA.  
XX  
AC AAG65781;  
XX  
DT 07-JAN-2002 (first entry)  
XX  
DE Amino acid sequence of HSPDB4A1-E222G fusion protein.  
XX  
KW PDE4; central nervous system; antiinflammatory; cytostatic; nootropic;  
KW autoimmune; ischemic; osteopathic; GFP; green fluorescent protein;  
KW fusion protein.  
XX  
OS Homo sapiens.  
OS Aequorea victoria.  
XX  
PN WO200179526-A2.  
XX  
PD 25-OCT-2001.  
XX  
PF 11-APR-2001; 2001WO-DK000264.  
XX  
PR 17-APR-2000; 2000DK-00000651.  
PR 29-MAY-2000; 2000DK-00000849.  
XX  
PA (BIOI-) BIOIMAGE AS.  
XX







CC biosensors of the invention can be used to investigate a wide range of  
CC cellular activities and to screen compounds which modulate these  
CC activities. Biosensors containing a recognition site for caspase, for  
CC example, may be used for the screening of compounds which modulate  
CC apoptosis, while biosensors containing other protease recognition sites  
CC may be used for the detection of proteolytic toxins (such as anthrax  
CC lethal factor). The method provides improved target validation and  
CC candidate compound optimisation by combining many cell screening formats  
CC with fluorescence-based molecular reagents and computer-based feature  
CC extraction, data analysis and automation, resulting in increased quantity  
CC and speed of data collection and faster evaluation of drug candidates.  
CC Sequences AAB2881-B2885 represent fluorescent proteins which may be used  
CC as components of biosensor fusion proteins of the invention  
XX  
SQ Sequence 239 AA;

Query Match 98.9%; Score 1260; DB 3; Length 239;  
Best Local Similarity 98.7%; Pred. No. 1.5e-121;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKLTLPKICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKLTLPKICTTGKLPVPWPT 60  
QY 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVOLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVOLA 180  
QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQALSCKDPNEKRDHVMVLFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGVPVLLPDNHYLSTQALSCKDPNEKRDHVMVLFVTAAGITLGMDELYK 239

RESULT 7  
AAV54349  
ID AAV54349 standard; protein; 239 AA.  
XX  
AC AAV54349;  
XX  
DT 06-APR-2000 (first entry)  
XX  
DE Amino acid sequence of the mutant green fluorescent protein EGFP.  
XX  
KW Fluorescent protein; green fluorescent protein; emission intensity;  
KW fluorescence; pH detection; pH sensor; EGFP.  
XX  
OS Synthetic.  
OS Aequorea victoria.  
XX

Key Location/Qualifiers  
FH Misc-difference 65  
FT /note= "wild type Phe substituted with Leu"  
FT Misc-difference 66  
FT /note= "wild type Ser substituted with Thr"  
FT Misc-difference 232  
FT /note= "wild type His substituted with Leu"  
XX  
XX WO964592-A2.  
XX  
XX 16-DEC-1999.  
XX  
XX 08-JUN-1999; 99WO-US012850.  
XX  
XX 09-JUN-1998; 98US-00094359.  
XX 13-OCT-1998; 98US-00172063.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
XX (UYOR-) UNIV OREGON STATE.  
XX

PI Tsien RY, Llopis J, Wachter RM;  
XX  
DR WPI; 2000-116540/10.  
DR N-PSDB; AAZ45642.  
XX  
PT New functional engineered green fluorescent proteins, used for measuring  
PT the pH in biological samples and cells.  
XX  
PS Disclosure; Page 9; 89pp; English.  
XX  
CC The present sequence represents a functional engineered fluorescent  
CC protein based on the Aequorea green fluorescent protein (GFP). The  
CC emission intensity changes as pH varies between 5 and 10 of the present  
CC protein are novel. The functional engineered fluorescent proteins show  
CC reversible changes in fluorescence over physiological pH ranges. They can  
CC be used for determining the pH of samples and cells. The polynucleotides  
CC can also be used to produce transgenic animals. The fluorescent protein  
CC pH sensors can be delivered to cells in the form of polynucleotides  
CC encoding the protein sensor fused to a targeting signal. The targeting  
CC signal directs the expression of the protein sensors to restricted cell  
CC locations. This makes it possible to measure the pH of a precisely  
CC defined cellular region or organelle  
XX  
SQ Sequence 239 AA;

Query Match 98.9%; Score 1260; DB 3; Length 239;  
Best Local Similarity 98.7%; Pred. No. 1.5e-121;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKLTLPKICTTGKLPVPWPT 60  
DB 1 MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYKLTLPKICTTGKLPVPWPT 60  
QY 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVOLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFIRHNIEDGSVOLA 180  
QY 181 DHYQONTPIGDGVPVLLPDNHYLSTQALSCKDPNEKRDHVMVLFVTAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGVPVLLPDNHYLSTQALSCKDPNEKRDHVMVLFVTAAGITLGMDELYK 239

RESULT 8  
AAV79584  
ID AAV79584 standard; peptide; 239 AA.  
XX  
AC AAV79584;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE EGFP signal domain.  
XX  
KW Protease; biosensor; EGFP; signal peptide; cell screening; assay;  
KW analysis; drug discovery.  
XX  
OS Unidentified.  
XX  
XX WO200026408-A2.  
XX  
XX 11-MAY-2000..  
XX  
XX 29-OCT-1999; 99WO-US025431.  
XX  
XX 30-OCT-1998; 98US-0106308P.  
XX 26-MAY-1999; 99US-0136078P.  
XX  
XX (CELL-) CELLOMICS INC.  
XX  
XX Guiliano KA, Bright G, Olson K, Burroughs-Tencza S;

XX WPI: 2000-365644/31.  
 DR N-PSDB; AA27573.  
 XX  
 PT Recombinant nucleic acid encoding a protease biosensor useful for  
 PT fluorescence based cell and molecular biochemical assays for drug  
 PT discovery comprising three operably linked nucleic acid sequences.  
 XX  
 XX Claim 14; Fig 29A; 218pp; English.  
 XX  
 CC The present sequence is that of the EGFP signal domain, which can be  
 CC included in novel recombinant protease biosensors (PBs) of the invention.  
 CC The PBs (see AAY79638-54) comprise: a first domain (see AAY79579-87)  
 CC comprising at least 1 detectable polypeptide signal such as the present  
 CC sequence; a second domain (see AAY79588-622) comprising at least 1  
 CC protease recognition site; and a third domain (see AAY79623-37)  
 CC comprising at least 1 reactant target sequence. A recombinant nucleic  
 CC acid (see AAA27627-43) encoding the PB, an expression vector, and a  
 CC genetically engineered host cell are also claimed. A claimed method for  
 CC identifying compounds that modify protease activity in a cell involves  
 CC contacting a host cell that possesses the recombinant PB with a test  
 CC compound, and determining the PB distribution in the host cell, where  
 CC changes in the distribution of the PB are correlated with modification of  
 CC protease activity by the test compound. Claimed kits for identifying  
 CC compounds that modify protease activity in a host cell include the  
 CC recombinant nucleic acid, or the recombinant PB, or the vector, or the  
 CC host cell. The PB is useful in high content screens to detect in vivo  
 CC activation of enzymatic activity, and to identify specific activity based  
 CC on cleavage of a known recognition motif  
 XX  
 SQ Sequence 239 AA;

Query Match 98.9%; Score 1260; DB 3; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 1.5e-121;  
 Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
 DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
 QY 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTTLTATYGVQCFSRYPDHMKQHDFFKSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNVMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 QY 181 DHYQQNTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGVFVTAAGITLGMDELYK 239  
 DB 181 DHYQQNTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGVFVTAAGITLGMDELYK 239

RESULT 9  
 AAB50804  
 ID AAB50804 standard; protein; 239 AA.  
 XX  
 AC AAB50804;  
 XX  
 DT 14-MAR-2001 (first entry)  
 XX  
 DE Jellyfish GFP mutant EGFP.  
 XX  
 KW Asquorea victoria; jellyfish; fluorescent protein indicator;  
 KW green fluorescent protein; GFP; linker moiety; sensor;  
 KW calmodulin-binding domain; mutant; mutein.  
 XX  
 OS Asquorea victoria.  
 XX  
 PN WO200071565-A2.  
 XX  
 PD 30-NOV-2000.  
 XX

PF 17-MAY-2000; 2000WO-US013684.  
 XX  
 PR 21-MAY-1999; 99US-00316919.  
 PR 21-MAY-1999; 99US-00316920.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Tsien RY, Baird GA;  
 XX  
 DR WPI; 2001-032017/04.  
 DR N-PSDB; AAC90488.  
 XX  
 PT Novel fluorescent proteins comprising a sensor protein inserted into  
 PT them, useful for measuring the response of a sensor biological, chemical,  
 PT electrical or physiological parameter in vivo or in vitro.  
 XX  
 XX Disclosure; Page 24; 94pp; English.  
 PS  
 CC The present sequence is a fluorescent protein used in the construction of  
 CC a fluorescent protein indicator. The indicator comprises a sensor  
 CC polypeptide that is responsive to a chemical, biological, electrical or  
 CC physiological parameter, and a fluorescence protein functional group. The  
 CC sensor polypeptide is operatively inserted into the fluorescent moiety.  
 CC The fluorescent indicator is useful for detecting the presence of a  
 CC response inducing member in a sample. The method involves contacting the  
 CC sample with the indicator and detecting a change in fluorescence, in  
 CC which a change is indicative of the effect of the parameter on the sensor  
 CC polypeptide. The novel fluorescent proteins are advantageous due to their  
 CC reduced size as compared to the FRET (fluorescence resonance energy  
 CC transfer)-based sensors  
 XX  
 SQ Sequence 239 AA;

Query Match 98.9%; Score 1260; DB 4; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 1.5e-121;  
 Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
 DB 1 MVSKEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGLTKLTKFICTTGKLPVWPWT 60  
 QY 61 LVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTTLTATYGVQCFSRYPDHMKQHDFFKSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNVMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNVMADKQNGIKVNFKIRHNIEDGVSQLA 180  
 QY 181 DHYQQNTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGVFVTAAGITLGMDELYK 239  
 DB 181 DHYQQNTPIGDGVPVLLPDNHYLSTQSALS KDPNEKRDHVMVLGVFVTAAGITLGMDELYK 239

RESULT 10  
 AAB85900  
 ID AAB85900 standard; protein; 239 AA.  
 XX  
 AC AAB85900;  
 XX  
 DT 30-NOV-2001 (first entry)  
 XX  
 DE A. victoria green fluorescent protein (GFP) and linker sequence.  
 XX  
 KW Melanin concentrating hormone receptor; MCHR; MCH; chimeric; fusion;  
 KW fluorescent polypeptide; orexinergic; anabolic; food intake; GFP;  
 KW green fluorescent protein.  
 XX  
 OS Synthetic.  
 OS Asquorea victoria.  
 XX  
 PN WO200168706-A1.  
 XX

PD 20-SEP-2001.  
 XX 14-MAR-2001; 2001WO-US008071.  
 PF  
 XX 15-MAR-2000; 2000US-0189698P.  
 PR  
 XX (MERI ) MERCK & CO INC.  
 PA  
 XX Marsh DJ;  
 PI  
 XX WPI; 2001-565791/63.  
 DR N-PSDB; AAH47304.  
 DR  
 XX Fusion proteins comprising melanin concentrating hormone receptor  
 PT peptides and fluorescent proteins, useful for identifying appetite  
 PT stimulants.  
 PT  
 XX Claim 2; Page 14; 71pp; English.  
 PS  
 XX The invention provides melanin concentrating hormone (MCH) receptor  
 CC (MCHR) chimeric and fusion proteins. The MCHR chimeric proteins comprise  
 CC MCHR polypeptide regions from different species. The MCHR fusion protein  
 CC comprise MCHR polypeptide region and a fluorescent polypeptide region  
 CC joined directly, or via a linker, to the carboxy side of the MCHR  
 CC polypeptide region. The MCHR fusion proteins can be expressed by standard  
 CC recombinant methodology. MCH action promotes feeding (orexigenic) and up  
 CC regulation of MCH activity stimulates food intake. The present sequence  
 CC represents a A. victoria green fluorescent protein (GFP) and a linker  
 CC sequence  
 XX  
 XX Sequence 239 AA;  
 SQ

Query Match 98.9%; Score 1260; DB 4; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 1.5e-121;  
 Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTLKFICTTGKLPVPWPT 60  
 DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTLKFICTTGKLPVPWPT 60  
 QY 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSGVOLA 180  
 DB 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSGVOLA 180  
 QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVGVFVTAAGITLGMDELYK 239  
 DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVGVFVTAAGITLGMDELYK 239

RESULT 11  
 AAB31171  
 ID AAB31171 standard; protein; 239 AA.  
 XX  
 AC AAB31171;  
 XX  
 XX 02-APR-2001 (first entry)  
 DT  
 XX Amino acid sequence of a green fluorescent protein (GFP).  
 DE  
 XX Growth rate; death rate; reporter gene; luminescent protein;  
 KW fluorescent product; luciferase; green fluorescent protein; GFP.  
 KW  
 XX Aequorea victoria.  
 OS  
 XX WO200075367-A1.  
 PN  
 XX 14-DEC-2000.  
 PD  
 XX 07-JUN-2000; 2000WO-FI000507.  
 PF

XX 07-JUN-1999; 99FI-00001296.  
 PR (LILI/) LILIUS E.  
 PA (VIRT/) VIRT M.  
 XX  
 PI Lilius E, Virta M;  
 XX WPI; 2001-061737/07.  
 DR N-PSDB; AAC86954.  
 DR  
 XX Assessing growth and death rates of a micro-organism in a desired  
 PT environment, by introducing 2 reporter genes encoding luminescent and  
 PT fluorescent products and detecting luminescent fluorescence.  
 XX  
 PS Disclosure; Page 27; 32pp; English.  
 PS  
 XX The specification describes a method for assessing the growth rate and  
 CC death rate of a micro-organism within a predetermined time period in a  
 CC desired environment. The method comprises introducing at least two  
 CC reporter genes encoding luminescent and/or fluorescent products into the  
 CC micro-organisms, incubating the micro-organism within the desired  
 CC environment, and detecting luminescence and/or fluorescence after a  
 CC predetermined time period. Use of two different markers within a micro-  
 CC organism enables the differentiation between growth and death rates. The  
 CC method is used to assess the growth rate and death rate of a micro-  
 CC organism within a predetermined time period in a desired environment. The  
 CC present sequence represents a green fluorescent protein (GFP), and is  
 CC encoded by a plasmid which encodes luminescent and fluorescent proteins,  
 CC and is used in the method of the invention  
 CC  
 XX Sequence 239 AA;  
 SQ

Query Match 98.9%; Score 1260; DB 4; Length 239;  
 Best Local Similarity 98.7%; Pred. No. 1.5e-121;  
 Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTLKFICTTGKLPVPWPT 60  
 DB 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGEDATYGLTLKFICTTGKLPVPWPT 60  
 QY 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 DB 61 LVTTLTYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120  
 QY 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSGVOLA 180  
 DB 121 VNRIELKGIDFEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSGVOLA 180  
 QY 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVGVFVTAAGITLGMDELYK 239  
 DB 181 DHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVGVFVTAAGITLGMDELYK 239

RESULT 12  
 AAG66198  
 ID AAG66198 standard; protein; 239 AA.  
 XX  
 AC AAG66198;  
 XX  
 XX 17-JUN-2002 (first entry)  
 DT  
 XX A. victoria green fluorescent protein (EGFP).  
 DE  
 XX Cyan-green fluorescent protein; fluorescence; recombinant; GFP;  
 KW green fluorescent protein; EGFP.  
 KW  
 XX Aequorea victoria.  
 OS  
 XX JP20002045189-A.  
 PN  
 XX 12-FEB-2002.  
 PD  
 XX



XX Key Location/Qualifiers  
FT Misc-difference 1. .3  
FT /note= "Wild-type GFP Met-Ser are replaced with Met-Val-Ser"  
FT  
FT Misc-difference 65  
FT /note= "GFP Phe64 is replaced by Leu"  
FT Misc-difference 66  
FT /note= "GFP Ser65 is replaced by Thr"  
XX  
XX  
PN EP1178109-A1.  
XX  
XX 06-FEB-2002.  
XX  
XX 03-AUG-2001; 2001EP-00306650.  
XX  
XX 04-AUG-2000; 2000JP-00237166.  
XX  
XX (RIKE ) RIKEN KK.  
XX  
XX Miyawaki A, Sawano A;  
XX WPI; 2002-208112/27.  
XX N-PSDB; AAD27910.  
XX  
XX Method for mutagenesis, e.g. for introducing certain or random mutations at certain sites of the nucleotide sequence, comprises synthesizing a mutated strand and a complementary strand by use of megaprimers.  
XX  
XX Example 1; Page 13-14; 31pp; English.  
XX  
XX The invention relates to a method for mutagenesis that comprises synthesizing a mutated strand and a complementary strand by use of megaprimers. The method basically comprises a DNA synthesis in which one or more primers that have a nucleotide sequence containing at least one mutation and a phosphorylated 5'-terminus are annealed to a template DNA and then subjected to an elongation reaction using a thermostable high-fidelity DNA polymerase, after which the phosphorylated 5'-terminus and the elongated terminus are ligated by means of a thermostable DNA ligase to synthesise a circular DNA containing the primers; a digestion in which the step of DNA synthesis is repeated several times to amplify the DNA containing the primers and then, at least DNAs other than the amplified circular DNA are digested into several fragments; and a double-stranded DNA synthesis in which, with the several fragments obtained in the step of digestion as megaprimers, the megaprimers are annealed to the circular DNA synthesised above, followed by an elongation reaction performed using the thermostable high-fidelity DNA polymerase. The method is useful for mutagenesis, particularly for introducing certain mutations at certain sites of a nucleotide sequence, or for introducing random mutations at certain sites of the nucleotide sequence. The present method is simple, speedy, economical and widely applicable. The present sequence is, Aequorea victoria enhanced green fluorescent protein (EGFP) used for mutagenesis in an exemplification of the invention. The EGFP is derived by altering the green fluorescent protein (GFP) sequence of Aequorea victoria  
XX  
SQ Sequence 239 AA;  
  
Query Match 98.9%; Score 1260; DB 5; Length 239;  
Best Local Similarity 98.7%; Pred. No. 1.5e-121;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 MVSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60  
DB 1 MVSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60  
  
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120  
DB 61 LVTTLYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120  
  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
  
QY 181 DHYQQNTPTICDGPVLLPDNHYLSTQSALSKDPNKRKDHMVLGVFTAAAGITLGMDELYK 239  
DB 181 DHYQQNTPTICDGPVLLPDNHYLSTQSALSKDPNKRKDHMVLGVFTAAAGITLGMDELYK 239

QY 181 DHYQQNTPTICDGPVLLPDNHYLSTQSALSKDPNKRKDHMVLGVFTAAAGITLGMDELYK 239  
DB 181 DHYQQNTPTICDGPVLLPDNHYLSTQSALSKDPNKRKDHMVLGVFTAAAGITLGMDELYK 239  
  
RESULT 15  
AAE34958  
ID AAE34958 standard; protein; 239 AA.  
XX  
XX AAE34958;  
XX  
XX 28-MAY-2003 (first entry)  
XX  
XX Aequorea victoria enhanced green fluorescent protein (EGFP).  
XX  
XX Phosphorylation indicator; fluorescent protein; detection; phosphatase;  
KW Kinase; enhanced green fluorescent protein; EGFP.  
XX  
XX Aequorea victoria.  
XX  
XX WO200295058-A2.  
XX  
XX 28-NOV-2002.  
XX  
XX 24-MAY-2002; 2002WO-US016955.  
XX  
XX 24-MAY-2001; 2001US-00865291.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
XX  
XX Tsien RY, Ting AY, Zhang J;  
XX WPI; 2003-148474/14.  
XX N-PSDB; AAD53428.  
XX  
XX Novel chimeric phosphorylation indicators, useful for detecting kinase/phosphatase in samples, has donor molecule, phosphorylatable domain, phosphaminoacid binding domain, and acceptor molecule, in operative linkage.  
XX  
XX Disclosure; Col 56-57; 38pp; English.  
XX  
XX The present invention relates to chimeric phosphorylation indicators comprising a phosphorylation polypeptide and a fluorescent protein or operative linkage, a donor molecule, a phosphorylatable domain, a phosphaminoacid binding domain (PAABD) and an acceptor molecule. The phosphorylation indicators of the invention are useful for detecting kinases or phosphatases in a biological sample. They are also useful in high throughput analysis e.g. for detecting a kinase inhibitor or phosphatase inhibitor. The present sequence is Aequorea victoria enhanced green fluorescent protein (EGFP) used in the invention  
XX  
SQ Sequence 239 AA;  
  
Query Match 98.9%; Score 1260; DB 6; Length 239;  
Best Local Similarity 98.7%; Pred. No. 1.5e-121;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 MVSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60  
DB 1 MVSKGELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTLCFKICTTGKLPVPWPT 60  
  
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120  
DB 61 LVTTLYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKTRAEVKFEGDTL 120  
  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
  
QY 181 DHYQQNTPTICDGPVLLPDNHYLSTQSALSKDPNKRKDHMVLGVFTAAAGITLGMDELYK 239  
DB 181 DHYQQNTPTICDGPVLLPDNHYLSTQSALSKDPNKRKDHMVLGVFTAAAGITLGMDELYK 239

Db 181 DHYQONTPIGDPVLLPDNNHYLSTQSALS KDPNEKRDMVLLFVTAAGITLGMDELYK 239

Search completed: June 21, 2004, 15:54:12  
Job time : 48.1111 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:55 ; Search time 12.7778 Seconds  
(without alignments)  
965.630 Million cell updates/sec

Title: US-09-887-784-222V

Perfect score: 1274

Sequence: 1 MVSKEELFTGVVPIVLVD.....VLGVFTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:\*
- 1: /cgn2\_6/prodata/2/iaa/5A COMB.pep.\*
  - 2: /cgn2\_6/prodata/2/iaa/5B COMB.pep.\*
  - 3: /cgn2\_6/prodata/2/iaa/6A COMB.pep.\*
  - 4: /cgn2\_6/prodata/2/iaa/6B COMB.pep.\*
  - 5: /cgn2\_6/prodata/2/iaa/PCTUS COMB.pep.\*
  - 6: /cgn2\_6/prodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1260	98.9	239	3	US-09-172-063-3
2	1260	98.9	239	4	US-09-513-783A-46
3	1260	98.9	239	4	US-09-316-919-4
4	1260	98.9	239	4	US-09-602-641-3
5	1260	98.9	239	4	US-09-920-922-2
6	1260	98.9	281	3	US-09-062-102-1
7	1260	98.9	281	4	US-09-364-946-1
8	1260	98.9	294	4	US-09-513-783A-2
9	1260	98.9	323	3	US-09-172-063-21
10	1260	98.9	323	4	US-09-602-641-21
11	1260	98.9	354	3	US-09-085-305-6
12	1260	98.9	379	4	US-09-417-197-129
13	1260	98.9	434	4	US-09-800-170-48
14	1260	98.9	442	4	US-09-417-197-127
15	1260	98.9	459	4	US-09-513-783A-170
16	1260	98.9	544	4	US-09-417-197-113
17	1260	98.9	544	4	US-09-417-197-115
18	1260	98.9	604	4	US-09-417-197-59
19	1260	98.9	605	4	US-09-417-197-41
20	1260	98.9	606	4	US-09-417-197-65
21	1260	98.9	607	4	US-09-417-197-47
22	1260	98.9	630	4	US-09-417-197-63
23	1260	98.9	631	4	US-09-417-197-39
24	1260	98.9	633	4	US-09-417-197-45
25	1260	98.9	635	4	US-09-417-197-125
26	1260	98.9	642	2	US-08-818-253-2
27	1260	98.9	642	2	US-08-818-253-6

ALIGNMENTS

RESULT 1

US-09-172-063-3  
; Sequence 3, Application US/09172063  
; Patent No. 6150176

; GENERAL INFORMATION:

; APPLICANT: Tsien, Roger Y.

; APPLICANT: Miyawaki, Atsushi

; APPLICANT: Liopis, Juan

; APPLICANT: Wachter, Rebekka M.

; APPLICANT: Remington, S. James

; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR

; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE

; FILE REFERENCE: 07257/071001

; CURRENT APPLICATION NUMBER: US/09/172,063

; CURRENT FILING DATE: 1998-10-13

; EARLIER APPLICATION NUMBER: 09/094,359

; EARLIER FILING DATE: 1998-06-09

; NUMBER OF SEQ ID NOS: 38

; SOFTWARE: FASTSEQ for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 239

; TYPE: PRT

; ORGANISM: Aequorea victoria

; FEATURE:

; NAME/KEY: VARIANT

; LOCATION: (0)...(0)

; OTHER INFORMATION: EGFP

US-09-172-063-3

Query Match 98.9%; Score 1260; DB 3; Length 239;

Best Local Similarity 98.7%; Pred. No. 1.5e-127;

Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY	1	MVSKGEELFTGVVPIVLVDGVNGHKFSVSGEGDATYVKLTAKFICTTGLKLPVDPWT	60
Db	1	MVSKGEELFTGVVPIVLVDGVNGHKFSVSGEGDATYVKLTAKFICTTGLKLPVDPWT	60
QY	61	LVTLLSYGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNVKTAEVKFEGDTL	120
Db	61	LVTLLTYGVQCFSRYPDHMKQHDFFKFSAMPEGVVQERTIFFKDDGNVKTAEVKFEGDTL	120
QY	121	VNRIELKIDIFKEDGNILGHKLEYNYSNNHYIMADKQNGIKVNFKIRHNIEDGSVQLA	180
Db	121	VNRIELKIDIFKEDGNILGHKLEYNYSNNHYIMADKQNGIKVNFKIRHNIEDGSVQLA	180
QY	181	DHYQONTPTIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGVFTAAAGITLGMDELYK	239
Db	181	DHYQONTPTIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGVFTAAAGITLGMDELYK	239

```
RESULT 2
US-09-513-783A-46
; Sequence 46, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 46
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGFP
US-09-513-783A-46

Query Match          98.9%; Score 1260; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239

RESULT 3
US-09-316-919-4
; Sequence 4, Application US/09316919
; Patent No. 6469154
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: 07257/073001
; CURRENT APPLICATION NUMBER: US/09/316,919
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-316-919-4

Query Match          98.9%; Score 1260; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239

RESULT 4
US-09-602-641-3
; Sequence 3, Application US/09602641
; Patent No. 6608189
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/602,641
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/172,063
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: EGFP
US-09-602-641-3

Query Match          98.9%; Score 1260; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239

RESULT 5
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. 6673610
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match          98.9%; Score 1260; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.5e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIILVELDGVNKGKFSVSGEGDATYVKLTFLKFICTTGKLPVPWPT 60
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVOLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239
```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match          98.9%; Score 1260; DB 4; Length 239;
Best Local Similarity 98.7%; Pred. No. 1.9e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEFCTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEFCTTGKLPVWPWT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKFSAMPEGVYVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKFSAMPEGVYVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239

RESULT 6
US-09-062-102-1
; Sequence 1, Application US/09062102
; Patent No. 6130313
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangqiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100
; CURRENT APPLICATION NUMBER: US/09/062,102
; CURRENT FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: US 60/060,855
; EARLIER FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6130313
US-09-062-102-1

Query Match          98.9%; Score 1260; DB 3; Length 281;
Best Local Similarity 98.7%; Pred. No. 1.9e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEFCTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEFCTTGKLPVWPWT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKFSAMPEGVYVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKFSAMPEGVYVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239
```

```

RESULT 7
US-09-364-946-1
; Sequence 1, Application US/09364946
; Patent No. 6306600
; GENERAL INFORMATION:
; APPLICANT: Kain, Steve
; APPLICANT: Li, Xiangqiang
; TITLE OF INVENTION: Rapidly Degrading GFP-Fusion Proteins and Methods
; FILE REFERENCE: D6100CIP/D2
; CURRENT APPLICATION NUMBER: US/09/364,946
; CURRENT FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: US 09/191,233
; EARLIER FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 281
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the EGFP-MODC422-461 fusion protein.
; Patent No. 6306600
US-09-364-946-1

Query Match          98.9%; Score 1260; DB 4; Length 281;
Best Local Similarity 98.7%; Pred. No. 1.9e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEFCTTGKLPVWPWT 60
DB 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEFCTTGKLPVWPWT 60
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKFSAMPEGVYVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKFSAMPEGVYVQERTIFFKDDGNGYKTRAEVKFEGDTL 120
QY 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNHNYSVIMADKQNGIKVNFKIRHNIEDGSVQLA 180
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239

RESULT 8
US-09-513-783A-2
; Sequence 2, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GFP-DEVD-Annexin II construct
US-09-513-783A-2

Query Match          98.9%; Score 1260; DB 4; Length 294;
Best Local Similarity 98.7%; Pred. No. 2.1e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTLEFCTTGKLPVWPWT 60
|||||
```

Db 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYGKLTLEFICTTGKLPVPWPT 60  
QY 61 LVTTLISGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 61 LVTTLISGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGVFTAAAGITLGMDELYK 239  
Db 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGVFTAAAGITLGMDELYK 239

RESULT 9

US-09-172-063-21  
; Sequence 21, Application US/09172063  
; Patent No. 6150176  
; GENERAL INFORMATION:  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Miyawaki, Atsushi  
; APPLICANT: Llopis, Juan  
; APPLICANT: Wachter, Rebekka M.  
; APPLICANT: Remington, S. James  
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
; FILE REFERENCE: 07257/071001  
; CURRENT APPLICATION NUMBER: US/09/172,063  
; CURRENT FILING DATE: 1998-10-13  
; EARLIER APPLICATION NUMBER: 09/094,359  
; EARLIER FILING DATE: 1998-06-09  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 21  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: GT-EGFP  
US-09-172-063-21

Query Match 98.9%; Score 1260; DB 3; Length 323;  
Best Local Similarity 98.7%; Pred. No. 2.4e-127;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYGKLTLEFICTTGKLPVPWPT 60  
Db 85 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYGKLTLEFICTTGKLPVPWPT 144  
QY 61 LVTTLISGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 145 LVTTLISGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 204  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
Db 205 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 264  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGVFTAAAGITLGMDELYK 239  
Db 265 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGVFTAAAGITLGMDELYK 323

RESULT 10

US-09-602-641-21  
; Sequence 21, Application US/09602641  
; Patent No. 6608189  
; GENERAL INFORMATION:  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Miyawaki, Atsushi  
; APPLICANT: Llopis, Juan

; APPLICANT: Wachter, Rebekka M.  
; APPLICANT: Remington, S. James  
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
; TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
; FILE REFERENCE: 07257/071001  
; CURRENT APPLICATION NUMBER: US/09/602,641  
; CURRENT FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 09/172,063  
; PRIOR FILING DATE: 1998-10-13  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 21  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: GT-EGFP  
US-09-602-641-21

Query Match 98.9%; Score 1260; DB 4; Length 323;  
Best Local Similarity 98.7%; Pred. No. 2.4e-127;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYGKLTLEFICTTGKLPVPWPT 60  
Db 85 MVSKEELFTGVVPILVELDGVNKGKFSVSGEGDATYGKLTLEFICTTGKLPVPWPT 144  
QY 61 LVTTLISGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 145 LVTTLISGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTL 204  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
Db 205 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 264  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGVFTAAAGITLGMDELYK 239  
Db 265 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLGVFTAAAGITLGMDELYK 323

RESULT 11

US-09-085-305-6  
; Sequence 6, Application US/09085305  
; Patent No. 6191269  
; GENERAL INFORMATION:  
; APPLICANT: Pollock, Allan  
; APPLICANT: Lovett, David H.  
; APPLICANT: Turck, Johanna  
; TITLE OF INVENTION: Selective Induction of Apoptosis in  
; TITLE OF INVENTION: Malignant Cancer Cells by Delivery of N-Terminal  
; TITLE OF INVENTION: Interleukin-1-Alpha Pro-Piece Polypeptide  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Bozicevic & Reed, LLP  
; STREET: 285 Hamilton Ave, Suite 200  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/085,305  
; FILING DATE: 29-MAY-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:

## ATTORNEY/AGENT INFORMATION:

NAME: Francis, Carol L  
REGISTRATION NUMBER: 36,513  
REFERENCE/DOCKET NUMBER: 6510/102US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-327-3400  
TELEFAX: 650-327-3231  
TELEX:

## INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:  
LENGTH: 364 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-09-085-305-6

Query Match 98.9%; Score 1260; DB 3; Length 364;  
Best Local Similarity 98.7%; Pred. No. 2.8e-127;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MYSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGKLTCLKFKICTTGKLPVPWPT 60  
DB 126 MYSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGKLTCLKFKICTTGKLPVPWPT 185  
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGVQERTIFFKDDGNGYKTRAEVKFEGDTL 120  
DB 186 LVTTLYGVQCFSRYPDHMKQHDFFKFSAMPEGVQERTIFFKDDGNGYKTRAEVKFEGDTL 245  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 246 VNRIELKGIDFKEDGNILGHKLEYNNSHNVMADKQNGIKVNFKIRHNIEDGSVQLA 305  
QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239  
DB 306 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 364

## RESULT 12

US-09-417-197-129  
Sequence 129, Application US/09417197  
Patent No. 6518021

## GENERAL INFORMATION:

APPLICANT: Ole THASTRUP, et al.  
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An I  
FILE REFERENCE: 3759-0110P  
CURRENT APPLICATION NUMBER: US/09/417,197  
CURRENT FILING DATE: 1999-10-07  
NUMBER OF SEQ ID NOS: 143  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 129  
LENGTH: 379  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: actin-binding-domain-EGFP fusion

US-09-417-197-129

Query Match 98.9%; Score 1260; DB 4; Length 379;  
Best Local Similarity 98.7%; Pred. No. 3e-127;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MYSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGKLTCLKFKICTTGKLPVPWPT 60  
DB 141 MYSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGKLTCLKFKICTTGKLPVPWPT 200  
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGVQERTIFFKDDGNGYKTRAEVKFEGDTL 120  
DB 201 LVTTLYGVQCFSRYPDHMKQHDFFKFSAMPEGVQERTIFFKDDGNGYKTRAEVKFEGDTL 260  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 261 VNRIELKGIDFKEDGNILGHKLEYNNSHNVMADKQNGIKVNFKIRHNIEDGSVQLA 320

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239  
DB 321 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 379

## RESULT 13

US-09-800-170-48  
Sequence 48, Application US/09800170  
Patent No. 6481667

## GENERAL INFORMATION:

APPLICANT: Kinsella, Todd  
TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES  
FILE REFERENCE: A-68614-1/DJB/RMS/RMK  
CURRENT APPLICATION NUMBER: US/09/800,170  
CURRENT FILING DATE: 2001-08-28  
PRIOR APPLICATION NUMBER: US 60/187,130  
PRIOR FILING DATE: 2000-03-06  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 48  
LENGTH: 434  
TYPE: PRT  
ORGANISM: Synechocystis PCC6803

US-09-800-170-48

Query Match 98.9%; Score 1260; DB 4; Length 434;  
Best Local Similarity 98.7%; Pred. No. 3.7e-127;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MYSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGKLTCLKFKICTTGKLPVPWPT 60  
DB 196 MYSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGKLTCLKFKICTTGKLPVPWPT 255  
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKFSAMPEGVQERTIFFKDDGNGYKTRAEVKFEGDTL 120  
DB 256 LVTTLYGVQCFSRYPDHMKQHDFFKFSAMPEGVQERTIFFKDDGNGYKTRAEVKFEGDTL 315

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 316 VNRIELKGIDFKEDGNILGHKLEYNNSHNVMADKQNGIKVNFKIRHNIEDGSVQLA 375

QY 181 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239  
DB 376 DHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 434

## RESULT 14

US-09-417-197-127  
Sequence 127, Application US/09417197  
Patent No. 6518021

## GENERAL INFORMATION:

APPLICANT: Ole THASTRUP, et al.  
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An Ir  
FILE REFERENCE: 3759-0110P  
CURRENT APPLICATION NUMBER: US/09/417,197  
CURRENT FILING DATE: 1999-10-07  
NUMBER OF SEQ ID NOS: 143  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 127  
LENGTH: 442  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: EGFP-RhoA fusion

US-09-417-197-127

Query Match 98.9%; Score 1260; DB 4; Length 442;  
Best Local Similarity 98.7%; Pred. No. 3.8e-127;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MYSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGKLTCLKFKICTTGKLPVPWPT 60

```
Db      1  MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
Qy      61  LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db      61  LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Qy     121  VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
Db     121  VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
Qy     181  DHYQQNTPIGDGPVLLPDNHYLSTQSALSADPNPKRDHVMVLGVFTAAGITLGMDELYK 239
Db     181  DHYQQNTPIGDGPVLLPDNHYLSTQSALSADPNPKRDHVMVLGVFTAAGITLGMDELYK 239
```

```
RESULT 15
US-09-513-783A-170
; Sequence 170, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 170
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP27
US-09-513-783A-170
```

```
Query Match      98.9%; Score 1260; DB 4; Length 459;
Best Local Similarity 98.7%; Pred. No. 4e-127;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1  MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
Db      1  MVSKEELFTGVVPILVELDGDVNGHKFVSQSGEGDATYGLTKLFICTTGKLPVPWPT 60
Qy     61  LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db     61  LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Qy     121  VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
Db     121  VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLA 180
Qy     181  DHYQQNTPIGDGPVLLPDNHYLSTQSALSADPNPKRDHVMVLGVFTAAGITLGMDELYK 239
Db     181  DHYQQNTPIGDGPVLLPDNHYLSTQSALSADPNPKRDHVMVLGVFTAAGITLGMDELYK 239
```

Search completed: June 21, 2004, 16:04:04  
Job time : 13.7778 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:54:20 ; Search time 34.7778 Seconds  
(without alignments)  
1940.117 Million cell updates/sec

Title: US-09-887-784-222V

Perfect score: 1274

Sequence: 1 MVSKEELFTGVVILVELD.....VLGVFTAAGITLGMDELYK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA:\*
- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
  - 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
  - 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
  - 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
  - 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
  - 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
  - 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
  - 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
  - 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
  - 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
  - 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
  - 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1271	99.8	239	9	US-09-887-784-4
2	1271	99.8	239	12	US-10-296-953-4
3	1271	99.8	363	14	US-10-270-223-6
4	1271	99.8	893	14	US-10-257-909A-30
5	1271	99.8	1132	14	US-10-257-909A-32
6	1263	99.1	239	9	US-09-887-784-2
7	1263	99.1	239	12	US-10-296-953-2
8	1260	98.9	239	9	US-09-920-922-2
9	1260	98.9	239	9	US-09-999-745-4
10	1260	98.9	239	10	US-09-866-538-4
11	1260	98.9	239	10	US-09-797-496B-2
12	1260	98.9	239	10	US-09-794-308-4
13	1260	98.9	239	10	US-09-865-291-4
14	1260	98.9	239	12	US-10-457-982-3
15	1260	98.9	239	14	US-10-121-258-13

16	1260	98.9	239	14	US-10-221-461-7	Sequence 7, Appli
17	1260	98.9	239	14	US-10-100-957A-46	Sequence 46, Appl
18	1260	98.9	239	14	US-10-177-390-2	Sequence 2, Appli
19	1260	98.9	239	14	US-10-338-411-3	Sequence 3, Appli
20	1260	98.9	239	15	US-10-370-570-4	Sequence 4, Appli
21	1260	98.9	239	15	US-10-389-640-3	Sequence 3, Appli
22	1260	98.9	259	14	US-10-314-861-11	Sequence 11, Appl
23	1260	98.9	281	12	US-09-931-232-1	Sequence 1, Appli
24	1260	98.9	288	14	US-10-314-861-37	Sequence 37, Appl
25	1260	98.9	293	14	US-10-314-861-35	Sequence 35, Appl
26	1260	98.9	294	14	US-10-100-957A-2	Sequence 2, Appli
27	1260	98.9	295	14	US-10-314-861-39	Sequence 39, Appl
28	1260	98.9	299	14	US-10-314-861-33	Sequence 33, Appl
29	1260	98.9	305	14	US-10-314-861-31	Sequence 31, Appl
30	1260	98.9	308	14	US-10-033-717-35	Sequence 35, Appl
31	1260	98.9	311	14	US-10-314-861-29	Sequence 29, Appl
32	1260	98.9	320	14	US-10-338-411-11	Sequence 11, Appl
33	1260	98.9	320	15	US-10-389-640-11	Sequence 11, Appl
34	1260	98.9	323	12	US-10-457-982-21	Sequence 21, Appl
35	1260	98.9	323	14	US-10-338-411-7	Sequence 7, Appli
36	1260	98.9	323	14	US-10-338-411-13	Sequence 13, Appl
37	1260	98.9	323	15	US-10-389-640-7	Sequence 7, Appli
38	1260	98.9	323	15	US-10-389-640-13	Sequence 13, Appl
39	1260	98.9	324	14	US-10-314-861-16	Sequence 16, Appl
40	1260	98.9	345	14	US-10-338-411-5	Sequence 5, Appli
41	1260	98.9	345	15	US-10-389-640-5	Sequence 5, Appli
42	1260	98.9	346	14	US-10-338-411-9	Sequence 9, Appli
43	1260	98.9	346	15	US-10-389-640-9	Sequence 9, Appli
44	1260	98.9	359	14	US-10-033-717-33	Sequence 33, Appl
45	1260	98.9	359	14	US-10-033-717-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1

US-09-887-784-4  
; Sequence 4, Application US/09887784  
; Patent No. US20020177189A1  
; GENERAL INFORMATION:  
; APPLICANT: RUORN, Sara et al  
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS  
; FILE REFERENCE: 3759-0115P  
; CURRENT APPLICATION NUMBER: US/09/887,784  
; CURRENT FILING DATE: 2001-06-19  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 4  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequoria Victoria  
US-09-887-784-4

Query Match 99.8%; Score 1271; DB 9; Length 239;  
Best Local Similarity 99.6%; Pred. No. 2.2e-124;  
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MVSKGEELFTGVVILVELDGVNGHKFSVSGEGDATYKLT	KFKICTTGKLPVPWPT	60
Db	1	MVSKGEELFTGVVILVELDGVNGHKFSVSGEGDATYKLT	KFKICTTGKLPVPWPT	60
Qy	61	LVTTLISYGVQCFSRYPDHMKQHDFFKSA	MPGVQVQRTIFFKDDGNVKTAEVKFEGDTL	120
Db	61	LVTTLISYGVQCFSRYPDHMKQHDFFKSA	MPGVQVQRTIFFKDDGNVKTAEVKFEGDTL	120
Qy	121	VNRIELKIDFKEDGNILGHKLEYNYSNHN	YIMADQKNGIKVNFKIRHNI	180
Db	121	VNRIELKIDFKEDGNILGHKLEYNYSNHN	YIMADQKNGIKVNFKIRHNI	180
Qy	181	DHYQONTPIGDGPVLLPDNHYLSTQSALS	KDPNEKRDHMLVGVFTAAGITLGMDELYK	239
Db	181	DHYQONTPIGDGPVLLPDNHYLSTQSALS	KDPNEKRDHMLVGVFTAAGITLGMDELYK	239

## RESULT 2

US-10-296-953-4  
; Sequence 4, Application US/10296953  
; Publication No. US20040072995A1  
; GENERAL INFORMATION:  
; APPLICANT: BJORN, SARA P.  
; APPLICANT: PAGLIARO, LEN  
; APPLICANT: THASTRUP, OLE  
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS  
; FILE REFERENCE: PL0095  
; CURRENT FILING DATE: 2002-11-26  
; PRIOR APPLICATION NUMBER: US/10/296,953  
; PRIOR FILING DATE: 2000-06-19  
; PRIOR APPLICATION NUMBER: PA 2000 00953  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: 60/212,681  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: 60/290,170  
; PRIOR FILING DATE: 2001-05-10  
; PRIOR APPLICATION NUMBER: PA 2001 00739  
; PRIOR FILING DATE: 2001-05-10  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
US-10-296-953-4

Query Match 99.8%; Score 1271; DB 12; Length 239;  
Best Local Similarity 99.6%; Pred. No. 2.2e-124;  
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MVSKGEELFTGVVPILVELDGDVNGHKFVS	VS	GE	GD	AT	Y	G	K	L	T	L	K	F	I	C	T	T	G	K	L	P	V	P	W	P	T	60							
Db	1	MVSKGEELFTGVVPILVELDGDVNGHKFVS	VS	GE	GD	AT	Y	G	K	L	T	L	K	F	I	C	T	T	G	K	L	P	V	P	W	P	T	60							
Qy	61	LVTTLISYGVQCFSRYPDHMKQHDFFKSAM	PE	GV	Q	VE	R	T	I	F	F	K	D	D	G	N	Y	K	T	R	A	E	V	K	F	E	G	D	T	L	120				
Db	61	LVTTLISYGVQCFSRYPDHMKQHDFFKSAM	PE	GV	Q	VE	R	T	I	F	F	K	D	D	G	N	Y	K	T	R	A	E	V	K	F	E	G	D	T	L	120				
Qy	121	VNRIELKGIDFKEDGNILGHKLEYNNSHN	VY	I	M	A	D	K	Q	N	G	I	K	V	N	F	K	I	R	H	N	I	E	D	G	S	V	Q	L	A	180				
Db	121	VNRIELKGIDFKEDGNILGHKLEYNNSHN	VY	I	M	A	D	K	Q	N	G	I	K	V	N	F	K	I	R	H	N	I	E	D	G	S	V	Q	L	A	180				
Qy	181	DHYQONTPIGDGPVLLPDNHYLSTQSA	L	S	K	D	P	N	E	K	R	D	H	M	V	L	V	G	F	V	T	A	A	G	I	T	L	G	M	D	E	L	Y	K	239
Db	181	DHYQONTPIGDGPVLLPDNHYLSTQSA	L	S	K	D	P	N	E	K	R	D	H	M	V	L	V	G	F	V	T	A	A	G	I	T	L	G	M	D	E	L	Y	K	239

## RESULT 3

US-10-270-223-6  
; Sequence 6, Application US/10270223  
; Publication No. US20030143634A1  
; GENERAL INFORMATION:  
; APPLICANT: Bioimage A/S  
; TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR COMPONENTS  
; TITLE OF INVENTION: INTERACT LIVING CELLS, AND TO EXTRACT QUANTITATIVE INFORMATION FROM THE SAME  
; FILE REFERENCE: 3759-0126P  
; CURRENT FILING DATE: 2002-10-11  
; CURRENT APPLICATION NUMBER: US/10/270,223  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 6  
; LENGTH: 363  
; TYPE: PRT  
; ORGANISM: Aequorea Victoria and Human  
US-10-270-223-6

Query Match 99.8%; Score 1271; DB 14; Length 363;  
Best Local Similarity 99.6%; Pred. No. 4.1e-124;  
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MVSKGEELFTGVVPILVELDGDVNGHKFVS	VS	GE	GD	AT	Y	G	K	L	T	L	K	F	I	C	T	T	G	K	L	P	V	P	W	P	T	60							
Db	1	MVSKGEELFTGVVPILVELDGDVNGHKFVS	VS	GE	GD	AT	Y	G	K	L	T	L	K	F	I	C	T	T	G	K	L	P	V	P	W	P	T	60							
Qy	61	LVTTLISYGVQCFSRYPDHMKQHDFFKSAM	PE	GV	Q	VE	R	T	I	F	F	K	D	D	G	N	Y	K	T	R	A	E	V	K	F	E	G	D	T	L	120				
Db	61	LVTTLISYGVQCFSRYPDHMKQHDFFKSAM	PE	GV	Q	VE	R	T	I	F	F	K	D	D	G	N	Y	K	T	R	A	E	V	K	F	E	G	D	T	L	120				
Qy	121	VNRIELKGIDFKEDGNILGHKLEYNNSHN	VY	I	M	A	D	K	Q	N	G	I	K	V	N	F	K	I	R	H	N	I	E	D	G	S	V	Q	L	A	180				
Db	121	VNRIELKGIDFKEDGNILGHKLEYNNSHN	VY	I	M	A	D	K	Q	N	G	I	K	V	N	F	K	I	R	H	N	I	E	D	G	S	V	Q	L	A	180				
Qy	181	DHYQONTPIGDGPVLLPDNHYLSTQSA	L	S	K	D	P	N	E	K	R	D	H	M	V	L	V	G	F	V	T	A	A	G	I	T	L	G	M	D	E	L	Y	K	239
Db	181	DHYQONTPIGDGPVLLPDNHYLSTQSA	L	S	K	D	P	N	E	K	R	D	H	M	V	L	V	G	F	V	T	A	A	G	I	T	L	G	M	D	E	L	Y	K	239

## RESULT 4

US-10-257-909A-30  
; Sequence 30, Application US/10257909A  
; Publication No. US20030187056A1  
; GENERAL INFORMATION:

; APPLICANT: Bernard R. TERRY et al.  
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular enzymes  
; FILE REFERENCE: 3759-0125P  
; CURRENT APPLICATION NUMBER: US/10/257,909A  
; CURRENT FILING DATE: 2002-10-17  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 30  
; LENGTH: 893  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:

; OTHER INFORMATION: Fusion between Aequorea victoria and human  
US-10-257-909A-30

Query Match 99.8%; Score 1271; DB 14; Length 893;  
Best Local Similarity 99.6%; Pred. No. 1.5e-123;  
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MVSKGEELFTGVVPILVELDGDVNGHKFVS	VS	GE	GD	AT	Y	G	K	L	T	L	K	F	I	C	T	T	G	K	L	P	V	P	W	P	T	60							
Db	655	MVSKGEELFTGVVPILVELDGDVNGHKFVS	VS	GE	GD	AT	Y	G	K	L	T	L	K	F	I	C	T	T	G	K	L	P	V	P	W	P	T	714							
Qy	61	LVTTLISYGVQCFSRYPDHMKQHDFFKSAM	PE	GV	Q	VE	R	T	I	F	F	K	D	D	G	N	Y	K	T	R	A	E	V	K	F	E	G	D	T	L	120				
Db	715	LVTTLISYGVQCFSRYPDHMKQHDFFKSAM	PE	GV	Q	VE	R	T	I	F	F	K	D	D	G	N	Y	K	T	R	A	E	V	K	F	E	G	D	T	L	774				
Qy	121	VNRIELKGIDFKEDGNILGHKLEYNNSHN	VY	I	M	A	D	K	Q	N	G	I	K	V	N	F	K	I	R	H	N	I	E	D	G	S	V	Q	L	A	180				
Db	775	VNRIELKGIDFKEDGNILGHKLEYNNSHN	VY	I	M	A	D	K	Q	N	G	I	K	V	N	F	K	I	R	H	N	I	E	D	G	S	V	Q	L	A	834				
Qy	181	DHYQONTPIGDGPVLLPDNHYLSTQSA	L	S	K	D	P	N	E	K	R	D	H	M	V	L	V	G	F	V	T	A	A	G	I	T	L	G	M	D	E	L	Y	K	239
Db	835	DHYQONTPIGDGPVLLPDNHYLSTQSA	L	S	K	D	P	N	E	K	R	D	H	M	V	L	V	G	F	V	T	A	A	G	I	T	L	G	M	D	E	L	Y	K	893

## RESULT 5

US-10-257-909A-32  
; Sequence 32, Application US/10257909A  
; Publication No. US20030187056A1  
; GENERAL INFORMATION:

; APPLICANT: Bernard R. TERRY et al.  
; TITLE OF INVENTION: Live cell procedures to identify compounds modulating intracellular enzymes  
; FILE REFERENCE: 3759-0125P  
; CURRENT APPLICATION NUMBER: US/10/257,909A  
; CURRENT FILING DATE: 2002-10-17  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 3.0



```

; SEQ ID NO 32
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion between Aequorea victoria and human
US-10-257-909A-32

Query Match
  Best Local Similarity 99.8%; Score 1271; DB 14; Length 1132;
  Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 60
DB 894 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 953

QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 954 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 1013

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFPIRHNIEDGVSQLA 180
DB 1014 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFPIRHNIEDGVSQLA 1073

QY 181 DHYQONTPIGDGPVLLPDNNHLYSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
DB 1074 DHYQONTPIGDGPVLLPDNNHLYSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 1132

RESULT 6
US-09-887-784-2
; Sequence 2, Application US/09887784
; Patent No. US20020177189A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, Sara et al
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: 3759-0115P
; CURRENT APPLICATION NUMBER: US/09/887,784
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-887-784-2

Query Match
  Best Local Similarity 99.1%; Score 1263; DB 9; Length 239;
  Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 60

QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFPIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFPIRHNIEDGVSQLA 180

QY 181 DHYQONTPIGDGPVLLPDNNHLYSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNNHLYSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239

RESULT 7
US-10-296-953-2
; Sequence 2, Application US/10296953
; Publication No. US2004007295A1
; GENERAL INFORMATION:
; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match
  Best Local Similarity 99.1%; Score 1263; DB 12; Length 239;
  Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 60

QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFPIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFPIRHNIEDGVSQLA 180

QY 181 DHYQONTPIGDGPVLLPDNNHLYSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNNHLYSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239

RESULT 8
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. US20020083489A1
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match
  Best Local Similarity 98.9%; Score 1360; DB 9; Length 239;
  Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 60
```

```

; APPLICANT: BJORN, SARA P.
; APPLICANT: PAGLIARO, LEN
; APPLICANT: THASTRUP, OLE
; TITLE OF INVENTION: NOVEL FLUORESCENT PROTEINS
; FILE REFERENCE: PL0095
; CURRENT APPLICATION NUMBER: US/10/296,953
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: PA 2000 00953
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 60/212,681
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/290,170
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PA 2001 00739
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-296-953-2

Query Match
  Best Local Similarity 99.2%; Score 1263; DB 12; Length 239;
  Matches 237; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 60

QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120
DB 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNYKTRAEVKFEGDTL 120

QY 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFPIRHNIEDGVSQLA 180
DB 121 VNRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQNGIKVNFPIRHNIEDGVSQLA 180

QY 181 DHYQONTPIGDGPVLLPDNNHLYSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239
DB 181 DHYQONTPIGDGPVLLPDNNHLYSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239

RESULT 8
US-09-920-922-2
; Sequence 2, Application US/09920922
; Patent No. US20020083489A1
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-2

Query Match
  Best Local Similarity 98.7%; Score 1360; DB 9; Length 239;
  Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 60
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFVSVEGEGDATYGKLTCLKFICTTGKLPVPWPT 60
```

QY 61 LVTTLTSGVOCFSGRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 61 LVTTLTSGVOCFSGRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHVQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239  
Db 181 DHVQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239  
RESULT 9  
US-09-999-745-4  
; Sequence 4, Application US/09999745  
; Patent No. US20020157120A1  
; GENERAL INFORMATION:  
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Baird, Geoffrey  
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS  
; FILE REFERENCE: REGEN1470-1  
; CURRENT APPLICATION NUMBER: US/09/999,745  
; CURRENT FILING DATE: 2001-10-23  
; PRIOR APPLICATION NUMBER: 09/316,920  
; PRIOR FILING DATE: 1999-05-21  
; NUMBER OF SEQ ID NOS: 67  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
US-09-999-745-4

Query Match 98.9%; Score 1260; DB 9; Length 239;  
Best Local Similarity 98.7%; Pred. No. 3.1e-123;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDATYVKLTLPICCTTGKLPVPWPT 60  
Db 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDATYVKLTLPICCTTGKLPVPWPT 60  
QY 61 LVTTLTSGVOCFSGRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 61 LVTTLTSGVOCFSGRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHVQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239  
Db 181 DHVQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239

RESULT 10  
US-09-866-538-4  
; Sequence 4, Application US/09866538  
; Publication No. US20030032088A1  
; GENERAL INFORMATION:  
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA  
; APPLICANT: Campbell, Robert  
; APPLICANT: Tsien, Roger  
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS  
; FILE REFERENCE: REGEN1530-2  
; CURRENT APPLICATION NUMBER: US/09/866,538  
; CURRENT FILING DATE: 2001-05-24  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 239  
; TYPE: PRT

; ORGANISM: Aequorea victoria  
US-09-866-538-4  
Query Match 98.9%; Score 1260; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 3.1e-123;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDATYVKLTLPICCTTGKLPVPWPT 60  
Db 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDATYVKLTLPICCTTGKLPVPWPT 60  
QY 61 LVTTLTSGVOCFSGRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 61 LVTTLTSGVOCFSGRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHVQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239  
Db 181 DHVQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239  
RESULT 11  
US-09-797-496B-2  
; Sequence 2, Application US/09797496B  
; Publication No. US20030049597A1  
; GENERAL INFORMATION:  
; APPLICANT: Simon, Sanford M.  
; APPLICANT: Chen, Yu  
; TITLE OF INVENTION: Chimeric Fluorescent Enzymes and Uses Thereof  
; FILE REFERENCE: 600-1-267  
; CURRENT APPLICATION NUMBER: US/09/797,496B  
; CURRENT FILING DATE: 2002-05-24  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Aequorea victoria green fluorescent protein modified as described  
US-09-797-496B-2

Query Match 98.9%; Score 1260; DB 10; Length 239;  
Best Local Similarity 98.7%; Pred. No. 3.1e-123;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDATYVKLTLPICCTTGKLPVPWPT 60  
Db 1 MVSKEELFTGVVPILVELDGVNKGHSVSGEGDATYVKLTLPICCTTGKLPVPWPT 60  
QY 61 LVTTLTSGVOCFSGRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTL 120  
Db 61 LVTTLTSGVOCFSGRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTL 120  
QY 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
Db 121 VNRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHVQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239  
Db 181 DHVQONTPIGDGPVLLPDNNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGMDELYK 239  
RESULT 12  
US-09-794-308-4  
; Sequence 4, Application US/09794308  
; Publication No. US20030170911A1  
; GENERAL INFORMATION:  
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA

APPLICANT: TSUEN, Roger  
APPLICANT: ZACHARIAS, David  
APPLICANT: BAIRD, Geoffrey  
TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS  
FILE REFERENCE: REG1530  
CURRENT APPLICATION NUMBER: US/09/794,308  
CURRENT FILING DATE: 2001-02-26  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
US-09-794-308-4

Query Match  
Best Local Similarity 98.9%; Score 1260; DB 10; Length 239;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVWPWT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVWPWT 60  
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQDPNEKRDHMLVGVFTAAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQDPNEKRDHMLVGVFTAAAGITLGMDELYK 239

RESULT 13  
US-09-865-291-4  
Sequence 4, Application US/09865291  
Publication No. US20030186229A1  
GENERAL INFORMATION:  
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA  
APPLICANT: TSUEN, Roger  
APPLICANT: TING, Alice  
APPLICANT: ZHANG, Jin  
TITLE OF INVENTION: EMISSION RATIONETRIC INDICATORS OF PHOSPHORYLATION  
FILE REFERENCE: REG1550  
CURRENT APPLICATION NUMBER: US/09/865,291  
CURRENT FILING DATE: 2001-05-24  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
US-09-865-291-4

Query Match  
Best Local Similarity 98.9%; Score 1260; DB 10; Length 239;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVWPWT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVWPWT 60  
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180

Query Match  
Best Local Similarity 98.7%; Pred. No. 3.1e-123;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVWPWT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVWPWT 60  
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180

QY 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQDPNEKRDHMLVGVFTAAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQDPNEKRDHMLVGVFTAAAGITLGMDELYK 239

RESULT 14  
US-10-457-982-3  
Sequence 3, Application US/10457982  
Publication No. US20030212265A1  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger Y.  
APPLICANT: Miyawaki, Atsushi  
APPLICANT: Liopis, Juan  
APPLICANT: Wachter, Rebekka M.  
APPLICANT: Remington, S. James  
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE  
FILE REFERENCE: 07257/071001  
CURRENT APPLICATION NUMBER: US/10/457,982  
CURRENT FILING DATE: 2003-06-09  
PRIOR APPLICATION NUMBER: US/09/602,641  
PRIOR FILING DATE: 2000-06-22  
PRIOR APPLICATION NUMBER: 09/172,063  
PRIOR FILING DATE: 1998-10-13  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Aequorea victoria  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (0)...(0)  
OTHER INFORMATION: EGFP  
US-10-457-982-3

Query Match  
Best Local Similarity 98.9%; Score 1260; DB 12; Length 239;  
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVWPWT 60  
DB 1 MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGDATYGLTKLFICTTGKLPVWPWT 60  
QY 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
DB 61 LVTTLSYGVCFSRYPDHMKQHDFFKSAPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTL 120  
QY 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180  
DB 121 VNRLEKGDIDFKEDGNILGHKLEYNHNHVMADKQNGIKVNFKIRHNIEDGSVQLA 180  
QY 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQDPNEKRDHMLVGVFTAAAGITLGMDELYK 239  
DB 181 DHYQONTPIGDGPVLLPDNHNHLSQTSALSQDPNEKRDHMLVGVFTAAAGITLGMDELYK 239

RESULT 15  
US-10-121-258-13  
Sequence 13, Application US/10121258  
Publication No. US20030059835A1  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger  
APPLICANT: Campbell, Robert  
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT  
TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME  
FILE REFERENCE: UC083.1CP2CP1  
CURRENT APPLICATION NUMBER: US/10/121,258  
CURRENT FILING DATE: 2002-04-10  
PRIOR APPLICATION NUMBER: 09/794,308  
PRIOR FILING DATE: 2001-02-26  
PRIOR APPLICATION NUMBER: 09/866,538  
PRIOR FILING DATE: 2001-05-24

```
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Enhanced Green Fluorescent Protein (EGFP)
US-10-121-258-13

Query Match      98.9%; Score 1260; DB 14; Length 239;
Best Local Similarity 98.7%; Pred. No. 3.1e-123;
Matches 236; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYKLTTLKFICTTGKLPVPWPT 60
Db      1 MVSKEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYKLTTLKFICTTGKLPVPWPT 60
Qy      61 LVTTLTYGVQCFSRYDPDMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Db      61 LVTTLTYGVQCFSRYDPDMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL 120
Qy      121 VNRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
Db      121 VNRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKVNFKIRHNIEDGSVOLA 180
Qy      181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239
Db      181 DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVGVFVTAAGITLGMDELYK 239
```

Search completed: June 21, 2004, 16:09:29  
Job time : 35.7778 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:46:00 ; Search time 10.3333 Seconds  
(without alignments)  
2224.817 Million cell updates/sec

Title: US-09-887-784-222V  
Perfect score: 1274  
Sequence: 1 MYSKGELFTGVVPIILVELD.....VLGVFVTAAGITLGMDELYK 239  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1238	97.2	238	1 JQ1514	green-fluorescent
2	104	8.2	785	2 H72228	hypothetical prote
3	91.5	7.2	861	2 H64102	leucine-trna ligas
4	90	7.1	632	2 T06586	DNA-binding protei
5	89.5	7.0	797	2 JC4078	protective surfac
6	89.5	7.0	808	2 F64102	protective surfac
7	89.5	7.0	887	2 E82590	leucyl-trna synthe
8	88.5	6.9	655	2 D83917	DNA topoisomerase
9	88.5	6.9	941	2 S29043	cellulase (EC 3.2.
10	88	6.9	578	1 I40794	dihydrolipoamide d
11	87.5	6.9	370	2 E70390	iron-sulfur cofact
12	87.5	6.9	860	2 AC0582	leucyl-trna synthe
13	87.5	6.9	2232	1 A36028	DNA-directed DNA p
14	87.5	6.9	2573	2 D71614	hypothetical prote
15	86	6.8	357	2 G81355	tRNA (uracil-5)-m
16	85.5	6.7	613	2 A99552	oligoendopeptidase
17	85.5	6.7	788	1 JDVLHH	DNA-directed DNA p
18	85.5	6.7	889	2 JC5576	inter-alpha-trypsi
19	85	6.7	281	2 AD2052	hypothetical prote
20	84.5	6.6	425	2 C97354	hypothetical prote
21	84	6.6	353	2 E84941	imidazoleglycerol-
22	84	6.6	874	2 JC4930	S-layer protein pr
23	83.5	6.6	836	1 JDVLJD	DNA-directed DNA p
24	83.5	6.6	1134	2 A60234	Iga Fc receptor pr
25	83.5	6.6	1164	1 FCSOAG	Iga Fc receptor pr
26	83	6.5	461	2 T06936	photosystem II chl
27	82.5	6.5	263	2 S53488	water-stress-induc
28	82.5	6.5	534	1 NICLMA	nitrogenase (EC 1.
29	82.5	6.5	740	2 G95153	neuraminidase, pro

ALIGNMENTS

RESULT 1

JQ1514  
green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)  
C:Species: Aequorea victoria  
C:Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 23-Mar-2001  
C:Accession: JS0692; JQ1514; PQ0335; S48693; S51330; S51331  
R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.  
Gene 111, 229-233, 1992  
A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.  
A:Reference number: JQ1514; MUID:92175527; PMID:1347277  
A:Accession: JS0692  
A:Molecule type: DNA  
A:Residues: 1-107, 'S', 109-238 <PEAL>  
A:Cross-references: GB:M62654; NID:G155662; PIDN:AAA27722.1; PID:G155663  
A:Accession: JQ1514  
A:Molecule type: mRNA  
A:Residues: 1-99, 'F', 'L', 101-140, 'L', 142-218, 'V', 220-238 <PRA2>  
A:Cross-references: GB:M62653; NID:G155660; PIDN:AAA27721.1; PID:G155661  
A:Accession: PQ0335  
A:Molecule type: protein  
A:Residues: 46-64; 74-122; 132-151; 154-183; 185-200 <PRA3>  
R:Inouye, S.; Tsuji, F.I.  
FEBS Lett. 351, 211-214, 1994  
A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.  
A:Reference number: S48693; MUID:94364470; PMID:8082767  
A:Accession: S48693  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-24, 'O', 26-156, 'P', 158-171, 'K', 173-238 <INO>  
A:Cross-references: GB:L29345; NID:G606383; PIDN:AAA58246.1; PID:G606384  
R:Watkins, J.N.; Campbell, A.K.  
submitted to the EMBL Data Library, January 1995  
A:Reference number: S51330  
A:Accession: S51330  
A:Molecule type: mRNA  
A:Residues: 1-13, 'V', 15-24, 'O', 26-44, 'N', 46-153, 'G', 155-156, 'P', 158-171, 'K', 173-227, 'R', 175-176, 'P', 177-178, 'K', 180-181, 'K', 182-183, 'K', 184-185, 'K', 186-187, 'K', 188-189, 'K', 190-191, 'K', 192-193, 'K', 194-195, 'K', 196-197, 'K', 198-199, 'K', 200-201, 'K', 202-203, 'K', 204-205, 'K', 206-207, 'K', 208-209, 'K', 210-211, 'K', 212-213, 'K', 214-215, 'K', 216-217, 'K', 218-219, 'K', 220-221, 'K', 222-223, 'K', 224-225, 'K', 226-227, 'K', 228-229, 'K', 230-231, 'K', 232-233, 'K', 234-235, 'K', 236-237, 'K', 238-239, 'K', 240-241, 'K', 242-243, 'K', 244-245, 'K', 246-247, 'K', 248-249, 'K', 250-251, 'K', 252-253, 'K', 254-255, 'K', 256-257, 'K', 258-259, 'K', 260-261, 'K', 262-263, 'K', 264-265, 'K', 266-267, 'K', 268-269, 'K', 270-271, 'K', 272-273, 'K', 274-275, 'K', 276-277, 'K', 278-279, 'K', 280-281, 'K', 282-283, 'K', 284-285, 'K', 286-287, 'K', 288-289, 'K', 290-291, 'K', 292-293, 'K', 294-295, 'K', 296-297, 'K', 298-299, 'K', 300-301, 'K', 302-303, 'K', 304-305, 'K', 306-307, 'K', 308-309, 'K', 310-311, 'K', 312-313, 'K', 314-315, 'K', 316-317, 'K', 318-319, 'K', 320-321, 'K', 322-323, 'K', 324-325, 'K', 326-327, 'K', 328-329, 'K', 330-331, 'K', 332-333, 'K', 334-335, 'K', 336-337, 'K', 338-339, 'K', 340-341, 'K', 342-343, 'K', 344-345, 'K', 346-347, 'K', 348-349, 'K', 350-351, 'K', 352-353, 'K', 354-355, 'K', 356-357, 'K', 358-359, 'K', 360-361, 'K', 362-363, 'K', 364-365, 'K', 366-367, 'K', 368-369, 'K', 370-371, 'K', 372-373, 'K', 374-375, 'K', 376-377, 'K', 378-379, 'K', 380-381, 'K', 382-383, 'K', 384-385, 'K', 386-387, 'K', 388-389, 'K', 390-391, 'K', 392-393, 'K', 394-395, 'K', 396-397, 'K', 398-399, 'K', 400-401, 'K', 402-403, 'K', 404-405, 'K', 406-407, 'K', 408-409, 'K', 410-411, 'K', 412-413, 'K', 414-415, 'K', 416-417, 'K', 418-419, 'K', 420-421, 'K', 422-423, 'K', 424-425, 'K', 426-427, 'K', 428-429, 'K', 430-431, 'K', 432-433, 'K', 434-435, 'K', 436-437, 'K', 438-439, 'K', 440-441, 'K', 442-443, 'K', 444-445, 'K', 446-447, 'K', 448-449, 'K', 450-451, 'K', 452-453, 'K', 454-455, 'K', 456-457, 'K', 458-459, 'K', 460-461, 'K', 462-463, 'K', 464-465, 'K', 466-467, 'K', 468-469, 'K', 470-471, 'K', 472-473, 'K', 474-475, 'K', 476-477, 'K', 478-479, 'K', 480-481, 'K', 482-483, 'K', 484-485, 'K', 486-487, 'K', 488-489, 'K', 490-491, 'K', 492-493, 'K', 494-495, 'K', 496-497, 'K', 498-499, 'K', 500-501, 'K', 502-503, 'K', 504-505, 'K', 506-507, 'K', 508-509, 'K', 510-511, 'K', 512-513, 'K', 514-515, 'K', 516-517, 'K', 518-519, 'K', 520-521, 'K', 522-523, 'K', 524-525, 'K', 526-527, 'K', 528-529, 'K', 530-531, 'K', 532-533, 'K', 534-535, 'K', 536-537, 'K', 538-539, 'K', 540-541, 'K', 542-543, 'K', 544-545, 'K', 546-547, 'K', 548-549, 'K', 550-551, 'K', 552-553, 'K', 554-555, 'K', 556-557, 'K', 558-559, 'K', 560-561, 'K', 562-563, 'K', 564-565, 'K', 566-567, 'K', 568-569, 'K', 570-571, 'K', 572-573, 'K', 574-575, 'K', 576-577, 'K', 578-579, 'K', 580-581, 'K', 582-583, 'K', 584-585, 'K', 586-587, 'K', 588-589, 'K', 590-591, 'K', 592-593, 'K', 594-595, 'K', 596-597, 'K', 598-599, 'K', 600-601, 'K', 602-603, 'K', 604-605, 'K', 606-607, 'K', 608-609, 'K', 610-611, 'K', 612-613, 'K', 614-615, 'K', 616-617, 'K', 618-619, 'K', 620-621, 'K', 622-623, 'K', 624-625, 'K', 626-627, 'K', 628-629, 'K', 630-631, 'K', 632-633, 'K', 634-635, 'K', 636-637, 'K', 638-639, 'K', 640-641, 'K', 642-643, 'K', 644-645, 'K', 646-647, 'K', 648-649, 'K', 650-651, 'K', 652-653, 'K', 654-655, 'K', 656-657, 'K', 658-659, 'K', 660-661, 'K', 662-663, 'K', 664-665, 'K', 666-667, 'K', 668-669, 'K', 670-671, 'K', 672-673, 'K', 674-675, 'K', 676-677, 'K', 678-679, 'K', 680-681, 'K', 682-683, 'K', 684-685, 'K', 686-687, 'K', 688-689, 'K', 690-691, 'K', 692-693, 'K', 694-695, 'K', 696-697, 'K', 698-699, 'K', 700-701, 'K', 702-703, 'K', 704-705, 'K', 706-707, 'K', 708-709, 'K', 710-711, 'K', 712-713, 'K', 714-715, 'K', 716-717, 'K', 718-719, 'K', 720-721, 'K', 722-723, 'K', 724-725, 'K', 726-727, 'K', 728-729, 'K', 730-731, 'K', 732-733, 'K', 734-735, 'K', 736-737, 'K', 738-739, 'K', 740-741, 'K', 742-743, 'K', 744-745, 'K', 746-747, 'K', 748-749, 'K', 750-751, 'K', 752-753, 'K', 754-755, 'K', 756-757, 'K', 758-759, 'K', 760-761, 'K', 762-763, 'K', 764-765, 'K', 766-767, 'K', 768-769, 'K', 770-771, 'K', 772-773, 'K', 774-775, 'K', 776-777, 'K', 778-779, 'K', 780-781, 'K', 782-783, 'K', 784-785, 'K', 786-787, 'K', 788-789, 'K', 790-791, 'K', 792-793, 'K', 794-795, 'K', 796-797, 'K', 798-799, 'K', 800-801, 'K', 802-803, 'K', 804-805, 'K', 806-807, 'K', 808-809, 'K', 810-811, 'K', 812-813, 'K', 814-815, 'K', 816-817, 'K', 818-819, 'K', 820-821, 'K', 822-823, 'K', 824-825, 'K', 826-827, 'K', 828-829, 'K', 830-831, 'K', 832-833, 'K', 834-835, 'K', 836-837, 'K', 838-839, 'K', 840-841, 'K', 842-843, 'K', 844-845, 'K', 846-847, 'K', 848-849, 'K', 850-851, 'K', 852-853, 'K', 854-855, 'K', 856-857, 'K', 858-859, 'K', 860-861, 'K', 862-863, 'K', 864-865, 'K', 866-867, 'K', 868-869, 'K', 870-871, 'K', 872-873, 'K', 874-875, 'K', 876-877, 'K', 878-879, 'K', 880-881, 'K', 882-883, 'K', 884-885, 'K', 886-887, 'K', 888-889, 'K', 890-891, 'K', 892-893, 'K', 894-895, 'K', 896-897, 'K', 898-899, 'K', 900-901, 'K', 902-903, 'K', 904-905, 'K', 906-907, 'K', 908-909, 'K', 910-911, 'K', 912-913, 'K', 914-915, 'K', 916-917, 'K', 918-919, 'K', 920-921, 'K', 922-923, 'K', 924-925, 'K', 926-927, 'K', 928-929, 'K', 930-931, 'K', 932-933, 'K', 934-935, 'K', 936-937, 'K', 938-939, 'K', 940-941, 'K', 942-943, 'K', 944-945, 'K', 946-947, 'K', 948-949, 'K', 950-951, 'K', 952-953, 'K', 954-955, 'K', 956-957, 'K', 958-959, 'K', 960-961, 'K', 962-963, 'K', 964-965, 'K', 966-967, 'K', 968-969, 'K', 970-971, 'K', 972-973, 'K', 974-975, 'K', 976-977, 'K', 978-979, 'K', 980-981, 'K', 982-983, 'K', 984-985, 'K', 986-987, 'K', 988-989, 'K', 990-991, 'K', 992-993, 'K', 994-995, 'K', 996-997, 'K', 998-999, 'K', 1000-1001, 'K', 1002-1003, 'K', 1004-1005, 'K', 1006-1007, 'K', 1008-1009, 'K', 1010-1011, 'K', 1012-1013, 'K', 1014-1015, 'K', 1016-1017, 'K', 1018-1019, 'K', 1020-1021, 'K', 1022-1023, 'K', 1024-1025, 'K', 1026-1027, 'K', 1028-1029, 'K', 1030-1031, 'K', 1032-1033, 'K', 1034-1035, 'K', 1036-1037, 'K', 1038-1039, 'K', 1040-1041, 'K', 1042-1043, 'K', 1044-1045, 'K', 1046-1047, 'K', 1048-1049, 'K', 1050-1051, 'K', 1052-1053, 'K', 1054-1055, 'K', 1056-1057, 'K', 1058-1059, 'K', 1060-1061, 'K', 1062-1063, 'K', 1064-1065, 'K', 1066-1067, 'K', 1068-1069, 'K', 1070-1071, 'K', 1072-1073, 'K', 1074-1075, 'K', 1076-1077, 'K', 1078-1079, 'K', 1080-1081, 'K', 1082-1083, 'K', 1084-1085, 'K', 1086-1087, 'K', 1088-1089, 'K', 1090-1091, 'K', 1092-1093, 'K', 1094-1095, 'K', 1096-1097, 'K', 1098-1099, 'K', 1100-1101, 'K', 1102-1103, 'K', 1104-1105, 'K', 1106-1107, 'K', 1108-1109, 'K', 1110-1111, 'K', 1112-1113, 'K', 1114-1115, 'K', 1116-1117, 'K', 1118-1119, 'K', 1120-1121, 'K', 1122-1123, 'K', 1124-1125, 'K', 1126-1127, 'K', 1128-1129, 'K', 1130-1131, 'K', 1132-1133, 'K', 1134-1135, 'K', 1136-1137, 'K', 1138-1139, 'K', 1140-1141, 'K', 1142-1143, 'K', 1144-1145, 'K', 1146-1147, 'K', 1148-1149, 'K', 1150-1151, 'K', 1152-1153, 'K', 1154-1155, 'K', 1156-1157, 'K', 1158-1159, 'K', 1160-1161, 'K', 1162-1163, 'K', 1164-1165, 'K', 1166-1167, 'K', 1168-1169, 'K', 1170-1171, 'K', 1172-1173, 'K', 1174-1175, 'K', 1176-1177, 'K', 1178-1179, 'K', 1180-1181, 'K', 1182-1183, 'K', 1184-1185, 'K', 1186-1187, 'K', 1188-1189, 'K', 1190-1191, 'K', 1192-1193, 'K', 1194-1195, 'K', 1196-1197, 'K', 1198-1199, 'K', 1200-1201, 'K', 1202-1203, 'K', 1204-1205, 'K', 1206-1207, 'K', 1208-1209, 'K', 1210-1211, 'K', 1212-1213, 'K', 1214-1215, 'K', 1216-1217, 'K', 1218-1219, 'K', 1220-1221, 'K', 1222-1223, 'K', 1224-1225, 'K', 1226-1227, 'K', 1228-1229, 'K', 1230-1231, 'K', 1232-1233, 'K', 1234-1235, 'K', 1236-1237, 'K', 1238-1239, 'K', 1240-1241, 'K', 1242-1243, 'K', 1244-1245, 'K', 1246-1247, 'K', 1248-1249, 'K', 1250-1251, 'K', 1252-1253, 'K', 1254-1255, 'K', 1256-1257, 'K', 1258-1259, 'K', 1260-1261, 'K', 1262-1263, 'K', 1264-1265, 'K', 1266-1267, 'K', 1268-1269, 'K', 1270-1271, 'K', 1272-1273, 'K', 1274-1275, 'K', 1276-1277, 'K', 1278-1279, 'K', 1280-1281, 'K', 1282-1283, 'K', 1284-1285, 'K', 1286-1287, 'K', 1288-1289, 'K', 1290-1291, 'K', 1292-1293, 'K', 1294-1295, 'K', 1296-1297, 'K', 1298-1299, 'K', 1300-1301, 'K', 1302-1303, 'K', 1304-1305, 'K', 1306-1307, 'K', 1308-1309, 'K', 1310-1311, 'K', 1312-1313, 'K', 1314-1315, 'K', 1316-1317, 'K', 1318-1319, 'K', 1320-1321, 'K', 1322-1323, 'K', 1324-1325, 'K', 1326-1327, 'K', 1328-1329, 'K', 1330-1331, 'K', 1332-1333, 'K', 1334-1335, 'K', 1336-1337, 'K', 1338-1339, 'K', 1340-1341, 'K', 1342-1343, 'K', 1344-1345, 'K', 1346-1347, 'K', 1348-1349, 'K', 1350-1351, 'K', 1352-1353, 'K', 1354-1355, 'K', 1356-1357, 'K', 1358-1359, 'K', 1360-1361, 'K', 1362-1363, 'K', 1364-1365, 'K', 1366-1367, 'K', 1368-1369, 'K', 1370-1371, 'K', 1372-1373, 'K', 1374-1375, 'K', 1376-1377, 'K', 1378-1379, 'K', 1380-1381, 'K', 1382-1383, 'K', 1384-1385, 'K', 1386-1387, 'K', 1388-1389, 'K', 1390-1391, 'K', 1392-1393, 'K', 1394-1395, 'K', 1396-1397, 'K', 1398-1399, 'K', 1400-1401, 'K', 1402-1403, 'K', 1404-1405, 'K', 1406-1407, 'K', 1408-1409, 'K', 1410-1411, 'K', 1412-1413, 'K', 1414-1415, 'K', 1416-1417, 'K', 1418-1419, 'K', 1420-1421, 'K', 1422-1423, 'K', 1424-1425, 'K', 1426-1427, 'K', 1428-1429, 'K', 1430-1431, 'K', 1432-1433, 'K', 1434-1435, 'K', 1436-1437, 'K', 1438-1439, 'K', 1440-1441, 'K', 1442-1443, 'K', 1444-1445, 'K', 1446-1447, 'K', 1448-1449, 'K', 1450-1451, 'K', 1452-1453, 'K', 1454-1455, 'K', 1456-1457, 'K', 1458-1459, 'K', 1460-1461, 'K', 1462-1463, 'K', 1464-1465, 'K', 1466-1467, 'K', 1468-1469, 'K', 1470-1471, 'K', 1472-1473, 'K', 1474-1475, 'K', 1476-1477, 'K', 1478-1479, 'K', 1480-1481, 'K', 1482-1483, 'K', 1484-1485, 'K', 1486-1487, 'K', 1488-1489, 'K', 1490-1491, 'K', 1492-1493, 'K', 1494-1495, 'K', 1496-1497, 'K', 1498-1499, 'K', 1500-1501, 'K', 1502-1503, 'K', 1504-1505, 'K', 1506-1507, 'K', 1508-1509, 'K', 1510-1511, 'K', 1512-1513, 'K', 1514-1515, 'K', 1516-1517, 'K', 1518-1519, 'K', 1520-1521, 'K', 1522-1523, 'K', 1524-1525, 'K', 1526-1527, 'K', 1528-1529, 'K', 1530-1531, 'K', 1532-1533, 'K', 1534-1535, 'K', 1536-1537, 'K', 1538-1539, 'K', 1540-1541, 'K', 1542-1543, 'K', 1544-1545, 'K', 1546-1547, 'K', 1548-1549, 'K', 1550



```
Db 409 PSKAKVNVLTQDLIVETFLRSVARTSGREGSEELKDSGNLERDTKKLELEQKNIS-E 467
Qy 126 LKGDIPKEDGNITGLHLE-----YNNSHNVYIMADKQK-----N 160
    |||: : : : : |||: : : : : |||: : : : :
Db 468 LKGIPTDNTLLDEKFEALGNKILKEISNPRHDVESANHSTHNKQVTVSHQKALETNN 527
Qy 161 GIKVNFKIRHNIEDG-----SVQLADHYQ 184
    : : : : : : : : : : : : : : : :
Db 528 QSQVEDVAKNIQDDSPSESLHKADKYR 557
    : : : : : : : : : : : : : : : :

RESULT 5
JC4078
protective surface antigen D-15 precursor - Haemophilus influenzae (type b)
C:Species: Haemophilus influenzae
A:Variety: type b
C:Date: 30-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 29-Sep-1999
C:Accession: JC4078
R:Flack, F.S.; Loosmore, S.; Chong, P.; Thomas, W.R.
Gene 156, 97-99, 1995
A:Title: The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus infl
A:Reference number: JC4078; MUID:95255676; PMID:7737523
A:Accession: JC4078
A:Molecule type: DNA
A:Residues: 1-797 <FLA>
A:Cross-references: GB:U13961; NID:G537447; PIDN:AAA85645.1; PID:G537448
A:Experimental source: type b
C:Superfamily: protective surface antigen D-15
C:Keywords: surface antigen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-797/Product: protective surface antigen D-15 #status predicted <MAT>

Query Match 7.0%; Score 89.5; DB 2; Length 797;
Best Local Similarity 22.4%; Pred. No. 18;
Matches 49; Conservative 29; Mismatches 78; Indels 63; Gaps 11;

Qy 65 LSYGVQCFRSYDPDHMKQHDF-----FKSAMPEGYVQE-----RTI 99
    : : : : : : : : : : : : : : : :
Db 427 IGYGTSGISYQASVKQDNFLGTGAASVIAGTKNDYGTSVNLGYTEPYFTKDGVS LGSNV 486
    : : : : : : : : : : : : : : : :

Qy 100 FPKDDGNKYTRAEVKFEGDTLVNRIELKGIDFKEDGNI---LGH-----KLEYNYS 148
    |||: : : : : |||: : : : : |||: : : : :
Db 487 PFENYDNKSDTSSNYKRTTYGSNVTL-GFPVNNNSYVGLGHTYKNISFALEYN--- 542
    : : : : : : : : : : : : : : : :

Qy 149 HNVYIMADKQK-NGIKVNFKIRHNIEDGVSQVLADHYQQ-----NTPIGDGPVLL 196
    : : : : : : : : : : : : : : : :
Db 543 RNLIIQSMKFKNGIKTN-----DFDFSFGWYNSLNRGYFTPKGVKASLG-GRVTI 593
    : : : : : : : : : : : : : : : :

Qy 197 P--DNHYLSTQSALS KDPNEKRDHMLVGVFTAAGITLG 233
    : : : : : : : : : : : : : : : :
Db 594 PGSDNKKYKLSADVQGFYPLDRDLRLWVWSAKASAGYANG 632
    : : : : : : : : : : : : : : : :

RESULT 6
F64102
protective surface antigen D-15 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 18-Sep-1998
C:Accession: F64102
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, J
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: F64102
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-808 <TIGR>
A:Cross-references: GB:I42023; TIGR:HI0917
C:Superfamily: protective surface antigen D-15
C:Keywords: surface antigen
```

```
Query Match 7.0%; Score 89.5; DB 2; Length 808;
Best Local Similarity 22.4%; Pred. No. 19;
Matches 49; Conservative 29; Mismatches 78; Indels 63; Gaps 11;

Qy 65 LSYGVQCFRSYDPDHMKQHDF-----FKSAMPEGYVQE-----RTI 99
    : : : : : : : : : : : : : : : :
Db 440 IGYGTSGISYQASVKQDNFLGTGAASVIAGTKNDYGTSVNLGYTEPYFTKDGVS LGSNV 499
    : : : : : : : : : : : : : : : :

Qy 100 FPKDDGNKYTRAEVKFEGDTLVNRIELKGIDFKEDGNI---LGH-----KLEYNYS 148
    |||: : : : : |||: : : : : |||: : : : :
Db 500 PFENYDNKSDTSSNYKRTTYGSNVTL-GFPVNNNSYVGLGHTYKNISFALEYN--- 555
    : : : : : : : : : : : : : : : :

Qy 149 HNVYIMADKQK-NGIKVNFKIRHNIEDGVSQVLADHYQQ-----NTPIGDGPVLL 196
    : : : : : : : : : : : : : : : :
Db 556 RNLIIQSMKFKNGIKTN-----DFDFSFGWYNSLNRGYFTPKGVKASLG-GRVTI 606
    : : : : : : : : : : : : : : : :

Qy 197 P--DNHYLSTQSALS KDPNEKRDHMLVGVFTAAGITLG 233
    : : : : : : : : : : : : : : : :
Db 607 PGSDNKKYKLSADVQGFYPLDRDLRLWVWSAKASAGYANG 645
    : : : : : : : : : : : : : : : :

RESULT 7
R82590
leucyl-tRNA synthetase XF2176 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: R82590
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: For a complete list of authors see reference number A59328 below
A:Accession: R82590
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-887 <SIM>
A:Cross-references: GB:AE004031; GB:AE003849; NID:g9107309; PIDN:AAF84975.1; GSPDB:GN0012
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; H
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
Submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Miyaki, C.Y.;
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tauhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF2176
C:Superfamily: leucine-tRNA ligase
```

```
Query Match 7.0%; Score 89.5; DB 2; Length 887;
Best Local Similarity 22.2%; Pred. No. 21;
Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;

Qy 50 TTGKLVPVPTLVTTLSYGVQCFRSYDPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNY-- 107
    : : : : : : : : : : : : : : : :
Db 329 TNEQLPV-WVAVFVLMAYGTGAVMVPVGHQDQDEE--ANKYGLPIQVIALKEPKNQDE 385
    : : : : : : : : : : : : : : : :

Qy 108 -----KTRAEVKFEGDTLVNRIELKGIDFKEDGNIHGKLEYNYSNHNVI 153
    |||: : : : : |||: : : : : |||: : : : :
Db 386 STWEPDVMRDVADKTR---EFE---LINSAFDGLDYQDAFEVLAERFE----- 429
    : : : : : : : : : : : : : : : :

Qy 154 MADKQKNG-IKVNFKIRHNIEDGVSQVLADHYQQNTPI-----GDGPPVLLPDN 199
    : : : : : : : : : : : : : : : :
Db 430 ---RQGRQRRVNYRLR-----DMGVSQRQYWGCPPIVYTCGAVPVPEQDLFVILPEN 482
    : : : : : : : : : : : : : : : :
```

```
QY 200 -HYLSTOSALSADPNKR 216
Db 483 VAFSGTGSPKTPDPEWRK 500

RESULT 8
D83917
DNA topoisomerase IV subunit B BH2140 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: D83917
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai, T.;
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: D83917
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-655 <STO>
A;Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BA05859.1; GSPDB:GN00
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH2140
C;Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B

Query Match 6.9%; Score 88.5; DB 2; Length 655;
Best Local Similarity 21.9%; Pred. No. 17;
Matches 53; Conservative 37; Mismatches 95; Indels 57; Gaps 9;

QY 22 DVNGHK--PVSVEGEGDAT---YGKLTLEFI-----CTTGKLPVWP 59
Db 63 NVTIHKQSVSDRDEGMPTGMHKLKPTPEVILTVLHAGGKFGGGYATSGLHGVA 122
QY 60 TLVTLTSLYGVQCFSRYPDPMKHQDFPKSAMPEGYVQER-----TIFPKDGG----- 105
Db 123 SVNALSEWLIVKIKRDGWVEQRFENGKPKSTLEKKGKTRGTGTHFKPDPTVFSTT 182
QY 106 --NYKTRAEVKFGDGLTVNRIELKGDGDFKEDGNILGHKLENNYNNHNYIMADK----- 157
Db 183 NFNVETLSERLRAAFLKGLKIELVDLRDDYTKVFH-YEDGKAFVYELNEDKETHPV 241
QY 158 -----QKNGIKVNFKIRHNIEDGSVOLADHYQONTPIGDQVLLPDNHYLSTOSALSADPN 212
Db 242 VPFNGSNGLEIEFAQFN--DGYTENVLSPVNNVTRKDG-----GTHELGAKTAMTRAV 294
QY 213 NE 214
Db 295 NE 296

RESULT 9
S29043
cellulase (EC 3.2.1.4) - Bacillus sp.
N;Alternate names: endo-1,4-beta-glucanase
C;Species: Bacillus sp.
C;Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 01-Dec-2000
C;Accession: S29043; PC4404
R;Ozaki, K.; Shikata, S.; Kawai, S.; Ito, S.; Okamoto, K.
J. Gen. Microbiol. 136, 1327-1334, 1990
A;Title: Molecular cloning and nucleotide sequence of a gene for alkaline cellulase from
A;Reference number: S29043; MUID:91037937; PMID:2230718
A;Accession: S29043
A;Molecule type: DNA
A;Residues: 1-941 <OZA>
A;Cross-references: EMBL:M27420; NID:g142664; PIDN:AAA22304.1; PID:g142665
R;Shirai, T.; Yamane, T.; Hidaka, T.; Kuyama, K.; Suzuki, A.; Ashida, T.; Ozaki, K.; Ito
J. Biochem. 122, 683-685, 1997
A;Title: Crystallization and preliminary X-ray analysis of a truncated family A alkaline
A;Reference number: PC4404; MUID:98060488; PMID:9399567
A;Accession: PC4404
A;Molecule type: protein
A;Residues: 228-584 <SHI>
A;Experimental source: strain KSM-635
```

```
C;Function:
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A;Pathway: cellulose degradation
C;Superfamily: Bacillus sp. KSM-635 alkaline cellulase; S-layer repeat homology; Thermotoc
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;41-95/Domain: S-layer repeat homology <SLR1>
F;101-153/Domain: S-layer repeat homology <SLR2>
F;164-219/Domain: S-layer repeat homology <SLR3>
F;766-908/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA>

Query Match 6.9%; Score 88.5; DB 2; Length 941;
Best Local Similarity 20.7%; Pred. No. 28;
Matches 50; Conservative 32; Mismatches 62; Indels 97; Gaps 11;

QY 16 LVELDGVNKHKPSVSGEGEGDATYVKLTLCITGKLPVWPPLVTLTSLYGVQCFSR 75
Db 240 LVELNG-----QUTLAGE---DGT-----PVQLRGMS---HGLQWFG-- 271
QY 76 PDHMKQHDFFKSAMPEGYVQERTIFPKDGNKYKTRAEVKFEGDGLVNRTELKGDIFKEDG 135
Db 272 -BIVNENAFVALSNDWGSNMIRLAMYIGENYATNPEVK---DLVVEGIELA----- 319
QY 136 NILGHKLENNYNNHNYIMADKQNGIKVNFIRHNIEDGSV-----QLADHYQONTPIGD 191
Db 320 -----FEHDMYIVDWH---VHAPGDPADVYSGAYDFFEEIADHYKDH----- 360
QY 192 GPVLLPDNHYLSTOSALSADPN-----EKRDHMLV 222
Db 361 -----PKNHYIITWELANEPSPNNNGPGLTNDEKGEAVKEYAEPIVEMLRKGDNMILV 415
QY 223 G 223
Db 416 G 416

RESULT 10
I40794
dihydrolipoamide dehydrogenase (EC 1.8.1.4) [validated] - Clostridium magnum
N;Alternate names: 2-oxoglutarate dehydrogenase complex chain E3; acetoin dehydrogenase
hydrogenase complex chain E3; S-complex 50K chain
C;Species: Clostridium magnum
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
A;Accession: I40794
R;Kruger, N.; Oppermann, F.B.; Lorenzl, H.; Steinbuechel, A.
J. Bacteriol. 176, 3614-3630, 1994
A;Title: Biochemical and molecular characterization of the Clostridium magnum acetoin de
A;Reference number: I40789; MUID:94266715; PMID:8206840
A;Accession: I40794
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-578 <KRU>
A;Cross-references: GB:L31844; NID:9472324; PIDN:AAA21748.1; PID:9472330
C;Function:
A;Description: catalyzes the oxidation of dihydrolipoamide to lipoamide using NAD
A;Pathway: acetoin dehydrogenase enzyme system
C;Superfamily: Alkaligenes dihydrolipoamide dehydrogenase; dihydrolipoamide dehydrogenase
C;Keywords: FAD; flavoprotein; lipoamide; NAD; oxidoreductase; redox-active disulfide
F;5-77/Domain: lipoyl/biotin-binding homology <LBP>
F;117-145/Region: beta-alpha-beta FAD nucleotide-binding fold
F;119-561/Domain: dihydrolipoamide dehydrogenase homology <DLID>
F;287-315/Region: beta-alpha-beta NAD nucleotide-binding fold
F;153-158/Disulfide bonds: redox-active #status predicted

Query Match 6.9%; Score 88; DB 1; Length 578;
Best Local Similarity 22.9%; Pred. No. 16;
Matches 57; Conservative 43; Mismatches 91; Indels 58; Gaps 13;

QY 10 TGWVILVELDGVNKHKPSVSGEGEGDATYVKLTLC-----FICTTGKLPVWPPLTV 63
Db 255 TGSMPFIPPIEIR---GNKLS-----GVIDST-GALLSLESNPESIAIGGVIGVFASIFN 305
QY 64 TLSYGVQCFSRYPDPMKHQHDFFKSAPEGYVQERTIFFKDDGNKYKTRAEVKFEGDGLTVNR 123
```



Db 306 SLGCKVSIIEMLPHILPPMDREISEI-----AKAKLIRDGININN 346

QY 124 IELKGIDFKEDG---NILGHKLEYNYNHNVMADKQKN--GIKVNFKIRHNIEDGSVQ 178

Db 347 CKVTRIEQEGDGLKVSFIGDKGESIDVEKVLIAVGRSNIIEGLDVE-KIGVKTEGGSII 405

QY 179 LADHYQONT-----PTGD--GPVLLPDNHYLSQTQSALSKD-----PNEKRDMHVMVLGVFV--- 225

Db 406 VNDKMETNVEGIYAIGDCTGKIMLA--HVASDQGVVAENIMGQNKMDYKTPACVYTK 463

QY 226 ---TAAGIT 231

Db 464 PELASVGLT 472

RESULT 11

E70390

iron-sulfur cofactor synthesis protein nifs - Aquifex aeolicus

N:Contains: L-cysteine sulfurtransferase (EC 2.8.1.1-)

C:Species: Aquifex aeolicus

C>Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 07-Dec-1999

C:Accession: E70390

R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: E70390

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-370 <AQF>

A:Cross-references: GB:AE000720; NID:g2983529; PIDN:AAC07111.1; PID:g2983536; GB:AE00065

A:Experimental source: strain VF5

C:Genetics:

A:Gene: nifs1

C:Superfamily: nitrogen fixation protein nifs

C:Keywords: phosphoprotein; pyridoxal phosphate; sulfurtransferase

F:195/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

F:318/Active site: Cys (cysteine persulfide intermediate) #status predicted

Query Match 6.9%; Score 87.5; DB 2; Length 370;

Best Local Similarity 25.4%; Pred. No. 9.9;

Matches 49; Conservative 30; Mismatches 91; Indels 23; Gaps 8;

QY 4 KGEELFTGVV---PILVELD---GDVNGHKF-SVSGEG---EGDATYVGLTKLFICT 50

Db 164 KGVPLLTDAVQAIKPIELKNIYSATFSGHKFHAIKGSGFLYISDEANYEFLIVCGGQE 223

QY 51 TGKLP-----VPVPTLVTTLSYGVQCFSRYPDHMKQ-HDFFKSAMPEGYVQERTIFFKDD 104

Db 224 NGRSGTENNVVGLSLAKALEIIVSNFSRYQEQKKLRDLFENLLLEA-LPDAQIVGKDA 282

QY 105 GNYKTRAEV---KFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNHNVMADKQKNG 161

Db 283 ERSPTSIVMPKFFGAIEIVNKLSEKGYICSTGSACLSGEYEPNKMHLKMGFSQEKALRM 342

QY 162 IKVNFKIRNIED 174

Db 343 VRFSFGLLNKEE 355

RESULT 12

AC0582

leucyl-tRNA synthetase [imported] - Salmonella enterica subsp. enterica serovar Typhi (S

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

C:Accession: AC0582

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AC0582

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-860 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD05125.1; PID:g16501899; GSPDB:GN00176

C:Genetics:

A:Gene: STY0699

C:Superfamily: leucine-tRNA ligase

Query Match 6.9%; Score 87.5; DB 2; Length 860;

Best Local Similarity 23.3%; Pred. No. 30;

Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;

QY 50 TTGKLVPWPPTLVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKT 109

Db 314 TGEETPV-WAANFVLMMEYGTGAVMAVPGH-DQRD-YEFASKYGLTIKPVILAADGSEPD 370

QY 110 RAEVKEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNHNVMADKQKNGIKVNFKIR 169

Db 371 SEQALTEKGVLFNSGFDGLAFEAFAFNAIADKL-----AEKGVGERKVNRYLR 418

QY 170 H-----NIEDGSVQLADHYQONTPIGDGPVLLPDNHYL-STQSALS 212

Db 419 DMGVSQRQYWGAPIPMVLTLEDGT-----LPTPDQLPVILPEVDVMDGITSPIKADP 471

RESULT 13

A36028

DNA-directed DNA polymerase (EC 2.7.7.7) II catalytic chain - yeast (Saccharomyces cerevi

N:Alternate names: DNA-directed DNA polymerase II chain A; protein N0825; protein YNL262v

C:Species: Saccharomyces cerevisiae

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000

C:Accession: A36028; B36028; S60919; S63235; S65121

R:Morrison, A.; Araki, H.; Clark, A.B.; Hamatake, R.K.; Sugino, A.

Cell 62, 1143-1151, 1990

A:Title: A third essential DNA polymerase in Saccharomyces cerevisiae.

A:Reference number: A36028; MUID:90381771; PMID:2169349

A:Accession: A36028

A:Molecule type: DNA

A:Residues: 1-2222 <MOR>

A:Cross-references: GB:M60416; GB:M36724; NID:g171408; PIDN:AAA88711.1; PID:g171409

A:Accession: B36028

A:Molecule type: protein

A:Residues: 1214-1216, 'X', 1218-1221 <MO2>

R:Sen-Gupta, M.; Lyck, R.; Fleig, U.; Niedenthal, R.N.; Hegemann, J.M.

submitted to the EMBL Data Library, October 1995

A:Description: The sequence of a 24152 bp segment from the left arm of chromosome XIV fr

A:Reference number: S60909

A:Accession: S60919

A:Molecule type: DNA

A:Residues: 1-2221 <SEN>

A:Cross-references: EMBL:X92494; NID:g1045236; PIDN:CAA63235.1; PID:g1045247

R:Sen-Gupta, M.; Lyck, R.; Niedenthal, R.K.; Fleig, U.N.; Hegemann, J.H.

submitted to the Protein Sequence Database, April 1996

A:Reference number: S63235

A:Accession: S63235

A:Molecule type: DNA

A:Residues: 1-2222 <SEW>

A:Cross-references: EMBL:271538; NID:g1302316; PIDN:CAA96169.1; PID:g1302317; GSPDB:GN000

A:Experimental source: strain S288C

R:Sen-Gupta, M.; Lyck, R.; Fleig, U.; Niedenthal, R.K.; Hegemann, J.H.

Yeast 12, 505-514, 1996

A:Title: The sequence of a 24 152 bp segment from the left arm of chromosome XIV from Sac

A:Reference number: S65111; MUID:96310631; PMID:8740425

A:Accession: S65121

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2221 <SEP>

A:Cross-references: EMBL:X92494; NID:g1045236; PIDN:CAA63235.1; PID:g1045247

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995

C:Genetics:

A;Gene: SGD:POL2; DUN2; MIPS:YNL262W  
A;Cross-references: SGD:S0005206; MIPS:YNL262W  
A;Map position: 14L  
C;Superfamily: DNA-directed DNA polymerase II  
C;Keywords: DNA binding; nucleotidyltransferase; nucleus; zinc finger

Query Match 6.9%; Score 87.5; DB 1; Length 2222;  
Best Local Similarity 28.2%; Pred. No. 1e+02;  
Matches 37; Conservative 14; Mismatches 49; Indels 31; Gaps 7;

QY 54 LPVPWP-TLVTLTSGVQCFSRVPDH-----KQHDFFKSAMPEGV-----QERTI 99  
Db 883 LPKSPFETFTLENGKLYLSYPCSMNLNVRVHQKFTNHQYQLKDLPLNVIYTHSENTI 942  
QY 100 FFKDDGNKYTR--AEVKFEGDTLVNR-----LELKGIDFKEDGNILGHKLEYNYN 147  
Db 943 FFEVDGPKAMILPSSKEEGKIKKRAYVFNEDGSLAELKGFELKRRGEL---QLIKNFQ 999  
QY 148 S--HNVYIMAD 156  
Db 1000 SDIFKVFLGCD 1010

RESULT 14  
D71614  
hypothetical protein PFB0460c - malaria parasite (Plasmodium falciparum)  
C;Species: Plasmodium falciparum  
C;Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
C;Accession: D71614  
R;Gardner, M.J.; Tetelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A;Reference number: A71600; MUID:99021743; PMID:9804551  
A;Accession: D71614  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-2573 <GAR>  
A;Cross-references: GB:AE001396; GB:AE001396; NID:g3845188; PIDN:AA71881.1; PID:g384519  
A;Experimental source: clone 3D7  
C;Genetics:  
A;Gene: PFB0460c

Query Match 6.9%; Score 87.5; DB 2; Length 2573;  
Best Local Similarity 26.2%; Pred. No. 1.2e+02;  
Matches 34; Conservative 30; Mismatches 53; Indels 13; Gaps 5;

QY 94 VQERTIFFKD--DGNKYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYN--YNSH 149  
Db 126 LKKEITILCKDIKSGNDPMDIEISLFKDDMVDDKELK--DPEKSKLKIKNKEVNFYNNKM 183  
QY 150 NVYIMADKQKNGIKVNFKIRHNIEDSGVOLADHYQQNTPIGDPGVLLPDNHYLSQTALS 209  
Db 184 NLHIKENKKKDEKKNKHNDNNNM-----IYYKNI---DKTHVILDNVNVHILNDIN 236  
QY 210 KDPNEKRDHM 219  
Db 237 TYLKRERDYM 246

RESULT 15  
G81355  
tRNA (uracil-5-)-methyltransferase (EC 2.1.1.35) Cj0831c [imported] - Campylobacter jejuni  
C;Species: Campylobacter jejuni  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
C;Accession: G81355  
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell  
Nature 403, 665-668, 2000  
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyf  
A;Reference number: A81250; MUID:20150912; PMID:10688204  
A;Accession: G81355  
A;Status: preliminary

A;Molecule type: DNA  
A;Residues: 1-357 <PAR>  
A;Cross-references: GB:ALJ39076; GB:ALJ111168; NID:g6968128; PIDN:CAB73096.1; PID:g696827;  
A;Experimental source: serotype O2, strain NCTC 11168  
C;Genetics:  
A;Gene: trmA; Cj0831c  
C;Keywords: methyltransferase; S-adenosylmethionine

Query Match 6.8%; Score 86; DB 2; Length 357;  
Best Local Similarity 24.8%; Pred. No. 13;  
Matches 30; Conservative 18; Mismatches 39; Indels 34; Gaps 5;

QY 80 KQHDFFKSAMPEGVYQERTIFFKDDGNKYKTRAEVKF--EGDTLV-----NRIELKG 128  
Db 14 EKHSFIKKYKFEFTYTKDFKLFASKDKKHVTRAELSFYHENDTLFYAMPDPKSKKKYIIIEY 73  
QY 129 IDPKED-----GNILGHKLEYNVNSHNYIMADKQKNGIKVNFKIRHNIE 173  
Db 74 LDFADEKICAFMPRLLEYLRLQDNKLKEKL-----FGVEFLTTKQE--LSITLLYHKNIE 125  
QY 174 D 174  
Db 126 D 126

Search completed: June 21, 2004, 16:01:59  
Job time : 11.3333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 15:42:24 ; Search time 6.44444 Seconds  
(without alignments)  
1931.085 Million cell updates/sec

Title: US-09-887-784-222V  
Perfect score: 1274  
Sequence: 1 MVSGBELFTGVVPLVELD.....VLGVFTAAIGTGMDELYK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1244	97.6	238	1 GFP_AEQVI	P42212 aequorea vi
2	91.5	7.2	861	1 SYL_HAEIN	P43827 haemophilus
3	89.5	7.0	795	1 D152_HAEIN	P44935 haemophilus
4	89.5	7.0	797	1 D151_HAEIN	P46024 haemophilus
5	89.5	7.0	879	1 SYL_XYLFA	Q9pbg8 xylella fas
6	88.5	6.9	793	1 D153_HAEIN	Q32629 haemophilus
7	88.5	6.9	941	1 GUN_EACS6	P19424 bacillus sp
8	87.5	6.9	860	1 SYL_SALTY	Q8z8h5 salmonella
9	87.5	6.9	860	1 SYL_SALTY	Q8zqz6 salmonella
10	87.5	6.9	2222	1 DPOE_YEAST	P21951 saccharomyc
11	87	6.8	689	1 AC2L_HUMAN	Q9nub1 homo sapien
12	86.5	6.8	533	1 CP51_CANGA	P50859 candida gla
13	86.5	6.8	879	1 SYL_XYLFT	Q87c65 xylella fas
14	86.5	6.8	1603	1 VIT4_CABEL	P18947 caenorhabdi
15	86	6.8	357	1 TRMA_CAMJE	Q9pp92 campylobact
16	85.5	6.7	613	1 PEPE_MYCPU	Q98dp0 mycoplasma
17	85.5	6.7	788	1 DPTH_HBHE	P13846 heron hepat
18	85.5	6.7	886	1 ITH3_MSAU	P97280 mesocricetu
19	85	6.7	504	1 YC03_KLEPN	Q48449 klebsiella
20	84.5	6.6	501	1 AMPA_WIGBR	Q8d295 wigleswort
21	84.5	6.6	859	1 SYL_SHRON	Q8ehp4 shewanella
22	84	6.6	353	1 HIS7_BUCAI	P57203 buchnera ap
23	84	6.6	366	1 SET7_HUMAN	Q8wt86 homo sapien
24	84	6.6	874	1 SLAP_BACLI	P49052 bacillus li
25	83.5	6.6	538	1 GRBE_RAT	Q88900 rattus norv
26	83.5	6.6	1164	1 BAG_STRAG	P27951 streptococc
27	83	6.5	461	1 PSBC_CVAPA	P48104 cyanophora
28	82.5	6.5	533	1 NIPD_CLOPA	P00467 clostridium
29	82	6.4	439	1 SY62_DLSOM	P24506 discopyge o
30	82	6.4	682	1 PRC_ECOLI	P23865 escherichia
31	82	6.4	752	1 NECI_RAT	P28840 rattus norv
32	82	6.4	774	1 AMY2_SCHPO	O42918 schizosacch
33	81.5	6.4	589	1 SYD_HABDU	Q7vnf0 haemophilus

RESULT 1						
GFP_AEQVI	ID	GFP_AEQVI	STANDARD;	PRT; 238 AA.		
AC	P42212;	Q17104; Q27903;				
DT	01-NOV-1995	(Rel. 32, Created)				
DT	01-NOV-1995	(Rel. 32, Last sequence update)				
DT	10-OCT-2003	(Rel. 42, Last annotation update)				
DE	Green fluorescent protein.					
GN	GFP.					
OS	Aequorea victoria (Jellyfish).					
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Hydromedusae; Aequoreidae; Aequorea.					
OX	NCBI_TaxID=6100;					
RN	[1]					
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.					
RX	MEDLINE=92175527;	PubMed=1347277;				
RA	Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,					
RA	Cormier M.J.;					
RT	"Primary structure of the Aequorea victoria green-fluorescent					
RT	protein."					
RL	Gene 111:229-233(1992).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=94185810;	PubMed=8137953;				
RA	Inouye S., Tsuji F.I.;					
RT	"Aequorea green fluorescent protein. Expression of the gene and					
RT	fluorescence characteristics of the recombinant protein.";					
RL	FEBS Lett. 341:277-280(1994).					
RN	[3]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=97299832;	PubMed=9154981;				
RA	Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;					
RT	"Enhanced expression in tobacco of the gene encoding green fluorescent					
RT	protein by modification of its codon usage.";					
RL	Plant Mol. Biol. 33:989-999(1997).					
RN	[4]					
RP	CHROMOPHORE.					
RX	MEDLINE=93192221;	PubMed=8448132;				
RA	Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;					
RT	"Chemical structure of the hexapeptide chromophore of the Aequorea					
RT	green-fluorescent protein.";					
RL	Biochemistry 32:1212-1218(1993).					
RN	[5]					
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).					
RX	MEDLINE=96355665;	PubMed=8703075;				
RA	Ormos M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,					
RA	Remington S.J.;					
RT	"Crystal structure of the Aequorea victoria green fluorescent					
RT	protein.";					
RL	Science 273:1392-1395(1996).					
RN	[6]					
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).					
RX	MEDLINE=98294543;	PubMed=9631087;				
RA	Yang F., Moss L.G., Phillips G.N. Jr.;					
RT	"The molecular structure of green fluorescent protein.";					
RL	Nat. Biotechnol. 14:1246-1251(1996).					

Q58743 methanococc  
Q99nb1 mus musculus  
P91679 drosophila  
Q25443 helicobacte  
O00116 homo sapien  
Q8xbn8 escherichia  
Q8fjy9 escherichia  
P07813 escherichia  
Q08694 chlamydomon  
Q9pk7 chlamydia m  
P94136 azorhizobiu  
P36924 bacillus ce

ALIGNMENTS

[7]  
X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.  
MEDLINE=98455509; PubMed=9782051;  
Wachter R.M., Eislinger M.A., Kallio K., Hanson G.T., Remington S.J.;  
"Structural basis of spectral shifts in the yellow-emission variants  
of green fluorescent protein.";  
Structure 6:1267-1277(1998).  
[8]  
X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
MEDLINE=99238303; PubMed=10220315;  
Eislinger M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J.;  
"Structural and spectral response of green fluorescent protein  
variants to changes in pH";  
Biochemistry 38:5296-5301(1999).  
-1- FUNCTION: Energy-transfer acceptor. Its role is to transduce the  
blue chemiluminescence of the protein aequorin into green  
fluorescent light by energy transfer. Fluoresces in vivo upon  
receiving energy from the Ca(2+)-activated photoprotein aequorin.  
Absorbs light maximally at 395 nm and exhibits a smaller  
absorbance peak at 470 nm. The fluorescence emission spectrum  
peaks at 509 nm with a shoulder at 540 nm.  
-1- SUBUNIT: Monomer.  
-1- TISSUE SPECIFICITY: Photocytes.  
-1- PTM: Contains a covalently attached chromophore, which is composed  
of modified amino acid residues. The chromophore is formed upon  
cyclization of the residues Ser-dehydroTy-Gly.  
-1- BIOTECHNOLOGY: Has become a useful and ubiquitous tool for making  
chimeric proteins of GFP linked to other proteins where it  
functions as a fluorescent protein tag. GFP tolerates N- and C-  
terminal fusion to a broad variety of proteins. It has been  
expressed in bacteria, yeast, slime mold, plants, Drosophila,  
zebrafish, and in mammalian cells. As a noninvasive fluorescent  
marker in living cells, it allows for a wide range of applications  
where it may function as a cell lineage tracer, reporter of gene  
expression, or as a measure of protein-protein interactions.  
-1- DATABASE: NAME-Protein Spotlight;  
NOTE-Issue 11 of June 2001;  
WWW="http://www.expasy.org/spotlight/articles/sptlt011.html".  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M62654; AAA27722.1; -;  
EMBL; M62653; AAA27721.1; -;  
EMBL; L29345; AAA58246.1; -;  
EMBL; X96418; CAA65278.1; -;  
PIR; JS0692; JQ1514.  
DR PDB; 1B9C; 17-NOV-00.  
DR PDB; 1BFP; 07-JUL-97.  
DR PDB; 1C4P; 14-JUN-00.  
DR PDB; 1EMA; 08-NOV-96.  
DR PDB; 1EMB; 16-JUN-97.  
DR PDB; 1EMC; 20-AUG-97.  
DR PDB; 1EMB; 20-AUG-97.  
DR PDB; 1EMF; 20-AUG-97.  
DR PDB; 1EMG; 12-MAY-99.  
DR PDB; 1EMK; 20-AUG-97.  
DR PDB; 1EML; 20-AUG-97.  
DR PDB; 1EMW; 20-AUG-97.  
DR PDB; 1F09; 17-NOV-00.  
DR PDB; 1F0B; 17-NOV-00.  
DR PDB; 1GFL; 11-JAN-97.  
DR PDB; 1HCJ; 15-JAN-97.  
DR PDB; 1HUY; 04-JUL-01.  
DR PDB; 1JBY; 07-JAN-03.  
DR PDB; 1JBZ; 07-JAN-03.  
DR PDB; 1KP5; 28-AUG-02.  
DR PDB; 1KVP; 10-APR-02.

DR PDB; 1KYR; 10-APR-02.  
DR PDB; 1KYS; 10-APR-02.  
DR PDB; 1YFP; 28-OCT-98.  
DR PDB; 2EMD; 20-AUG-97.  
DR PDB; 2EMN; 20-AUG-97.  
DR PDB; 2EMO; 20-AUG-97.  
DR InterPro; IPR009017; GFP like.  
DR InterPro; IPR000786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFP; 1.  
DR PRODOM; PD013756; Green\_fl\_protein; 1.  
KW Luminescence; 3D-structure.  
FT CROSSLINK 65 67 5-imidazolinone (Ser-Gly).  
FT MOD\_RES 66 66 2,3-DIDEHYDROTYROSINE.  
FT VARIANT 100 100 F -> Y.  
FT VARIANT 108 108 T -> S.  
FT VARIANT 141 141 L -> M.  
FT VARIANT 219 219 V -> I.  
FT CONFLICT 2 2 S -> G (IN REF. 3).  
FT CONFLICT 25 25 H -> Q (IN REF. 2).  
FT CONFLICT 80 80 Q -> R (IN REF. 3).  
FT CONFLICT 157 157 Q -> P (IN REF. 2).  
FT CONFLICT 172 172 E -> K (IN REF. 2).  
FT HELIX 4 8  
FT STRAND 12 22  
FT TURN 23 24  
FT STRAND 25 36  
FT TURN 37 40  
FT STRAND 41 48  
FT TURN 49 50  
FT HELIX 57 60  
FT TURN 61 63  
FT HELIX 69 71  
FT STRAND 73 73  
FT HELIX 76 81  
FT HELIX 83 86  
FT TURN 87 90  
FT STRAND 92 100  
FT TURN 101 102  
FT STRAND 105 115  
FT TURN 116 117  
FT STRAND 118 128  
FT TURN 132 133  
FT TURN 135 139  
FT STRAND 141 141  
FT STRAND 148 155  
FT TURN 156 159  
FT STRAND 160 171  
FT TURN 172 173  
FT STRAND 176 187  
FT STRAND 199 208  
FT TURN 211 212  
FT STRAND 217 227  
SQ SEQUENCE 238 AA; 26886 MW; EASA6F21FBFB6E05 CRC64;  
Query Match 97.6%; Score 1244; DB 1; Length 238;  
Best Local Similarity 97.9%; Pred. No. 3 1e-95;  
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
QY 2 VSKGEELFTGWVPIVLVELDGVNNGHKFSVSGEGDATTGKLTGKLTCTTGKLPVPTL 61  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 1 MSKGEELFTGWVPIVLVELDGVNNGHKFSVSGEGDATTGKLTGKLTCTTGKLPVPTL 60  
QY 62 VTTLVSVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIPKDDGNKTRAEVKEFGDTLV 121  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIPKDDGNKTRAEVKEFGDTLV 120  
QY 122 NRTELKGDIFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLAD 181  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 121 NRTELKGDIFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLAD 180  
QY 182 HYQONTPIGDGPVLLPDNHYLSTQSSALSXDPNKRDMVLVGFVTAAGITLGHDELYK 239  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

Db 181 HYQNTPIGDGPVLLPDNHYLSQSALSQKDPNEKRDHVMLEFVTAAGITHGMDELYK 238
RESULT 2
SYL_HAEIN
ID SYL_HAEIN STANDARD; PRT; 861 AA.
AC P43827;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR HI0921.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Rd / KW20 / ATCC 51907;
RC MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32774; AAC22581.1; -
CC PIR; H64102; H64102.
CC TIGR; HI0921; -
CC HAMAP; MF 00049; -; 1.
CC InterPro; IPR002302; Leu-tRNA-synt1a.
CC InterPro; IPR002300; tRNA-synt_1a.
CC InterPro; IPR001412; tRNA-synt_1.
CC InterPro; IPR009008; ValRS_1lers_edit.
CC Pfam; PF00133; tRNA-synt 1; 1.
CC PRINTS; P00985; TRNASYNTHLEU.
CC TIGRFAMs; TIGR00396; leuS_bact; 1.
CC PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
SQ SEQUENCE 861 AA; 97750 MW; EB93304F6B4C8FE7 CRC64;
Query Match 7.2%; Score 91.5; DB 1; Length 861;
Best Local Similarity 24.1%; Pred. No. 6.9;
Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;
QY 50 TTGKLPVNPFTLVTLLSYGVQCFSRYPDHMKQHDFFKSAWPEGYQVETIFFKQD-----103
Db 314 TGDKLPV-NVANFVLMHYGTGAVMAVPAH-DQRDF-----EFAQKYSLPKQVIAPLA 364

```

---

```

QY 104 DGNKYTRAEVKEGDTLVNRIELKGIDFKEDGNILGHKLEYNHNYSNVIMADK-QKNGI 162
Db 365 DEEIDLTKQAFVEHGHKLVNSDFDGKNF--DGAENG-----IADKLEKLG 408
QY 163 ----KVNFKIRH-----NIEDGSVQLADHYQYQNTPIGDGPVLLPDNHYL- 202
Db 409 GKRQVNYRLRDWGVSRQRYWGAPIPMLTLENGDVVPA-----PMEDLPILPEDVYMD 461
QY 203 STQSALSQKDPN 213
Db 462 GVKSPINADPN 472
RESULT 3
D152_HAEIN
ID D152_HAEIN STANDARD; PRT; 795 AA.
AC P44935;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
GN HI0917.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Rd / KW20 / ATCC 51907;
RC MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: Belongs to the surface antigen D15 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32773; AAC22575.1; -
CC TIGR; HI0917; -
CC InterPro; IPR000184; Bac_surfAg_D15.
CC Pfam; PF01103; Bac_surfaceAg; 1.
KW Antigen; Outer membrane; Signal; Complete proteome.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 795 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 795 AA; 87478 MW; B85691FC22B6ED44 CRC64;
Query Match 7.0%; Score 89.5; DB 1; Length 795;
Best Local Similarity 22.4%; Pred. No. 9.2;
Matches 49; Conservative 29; Mismatches 78; Indels 63; Gaps 11;
QY 65 LSYGVQCFSRYPDHMKQHDFF-----FKSAMPEGYVQE-----RTI 99
Db 427 IGVGTESGYSYQASVQKDNFLGTGAASVAGTAKNDYGTSLNLTGTFPYFTKDGVS LGNV 486
QY 100 FFKDDGNYKTRAEVKEGDTLVNRIELKGIDFKEDGNI---LGH-----KLEYNYS 148

```

```
Db 487 FFENYDMSKSDTSSNYKRTTYGNSVTL-GFPVNNENSYVGLGHTYNNKISNFALEYN--- 542
Qy 149 HNVYIMADKQK-NGIKVNFKIRHNIEDGSGVOLADHYQQ-----NTPIGDGPVLL 196
Db 543 RNLXIQSMKFKGNGIKTN-----DFDFSFGWNNYNSLNRGYFPYTKGVKASLG-GRVTI 593
Qy 197 P--DNHYLSTQSALSKDPNEKRDHVMVLGVFTAAAGITLG 233
Db 594 PGSDNKYYKLSADVQGFYPLDRDLHLWVVSASAGSAGYANG 632

RESULT 4
ID D151_HAEIN STANDARD; PRT; 797 AA.
AC P46024;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) Outer membrane protein D15.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype B;
RX MEDLINE=95255676; PubMed=7737523;
RA Flack F.S., Loosmore S., Chong P., Thomas W.R.;
RT "The sequencing of the 80-kDa D15 protective surface antigen of
RL Haemophilus influenzae.";
RL Gene 156:97-99(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Minna / Serotype B, and Eagan / Serotype B;
RX MEDLINE=97427952; PubMed=9284140;
RA Loosmore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease.";
RL Infect. Immun. 65:3701-3707(1997).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: Belongs to the surface antigen D15 family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U13961; AAA85645.1; -
CC DR EMBL; U60832; AAB61974.1; -
CC DR EMBL; U60833; AAB61976.1; -
CC DR PIR; JC4078; JC4078.
CC DR InterPro; IPR000184; Bac_surfAg_D15.
CC DR Pfam; PF01103; Bac_surface_Ag; I.
CC KW Antigen; Outer membrane; Signal.
CC FT SIGNAL 1 19 POTENTIAL.
CC FT CHAIN 20 797 PROTECTIVE SURFACE ANTIGEN D15.
CC SEQUENCE 797 AA; 87675 MW; 2F93DE538696AF1B CRC64;

Query Match 7.0%; Score 89.5; DB 1; Length 797;
Best Local Similarity 22.4%; Pred. No. 9.2;
Matches 49; Conservative 29; Mismatches 63; Gaps 11;

Qy 65 LSYGVQCFSPYDPHMKQHDH-----FKSAMPEGYVOE-----RTI 99
Db 427 IGVGTSGISYQASVKQDNFLGTGAASVAGTAKNDYGTNSVNLGYTPYFTKDGVSIGGNV 486
Qy 100 FFKDDGNKYTRAEVKFEGDTLVNRIELKGIDFKEDGNI---LGH-----KLEYNYS 148
```

```
Db 487 FFENYDMSKSDTSSNYKRTTYGNSVTL-GFPVNNENSYVGLGHTYNNKISNFALEYN--- 542
Qy 149 HNVYIMADKQK-NGIKVNFKIRHNIEDGSGVOLADHYQQ-----NTPIGDGPVLL 196
Db 543 RNLXIQSMKFKGNGIKTN-----DFDFSFGWNNYNSLNRGYFPYTKGVKASLG-GRVTI 593
Qy 197 P--DNHYLSTQSALSKDPNEKRDHVMVLGVFTAAAGITLG 233
Db 594 PGSDNKYYKLSADVQGFYPLDRDLHLWVVSASAGSAGYANG 632

RESULT 5
ID SYL_XYLFA STANDARD; PRT; 879 AA.
AC Q9P6G8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR XF2176.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5G; PubMed=10910347;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Cartaro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.B., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE004031; AAF84975.1; ALT_INIT.
CC DR HAMAP; MF_00049; -; 1.
CC InterPro; IPR020302; Leu-tRNA-synt1a.
CC
```

```
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR009008; Valrsls_edit.
DR Pfam; PF00133; tRNA-synt_1.1.
DR PRINTS; PR00985; TRNASYNTHLSU.
DR TIGRPFAM; TIGR00396; leuS_bact; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 45 55 "HIGH" REGION.
FT SITE 637 641 "RMSKS" REGION.
FT BINDING 640 640 ATP (BY SIMILARITY).
SQ SEQUENCE 879 AA; 99796 MW; 9FDCB992092919E CRC64;

Query Match
Best Local Similarity 22.2%; Pred. No. 10;
Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;

QY 50 TTGKLPVPMPTLVTLLSYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFPKDQGNV-- 107
DB 321 TNEQLPV-VVANFVLMAYGTGAVMVPGHQDRQDEP--ANKYGLPIRQVIALKEPKNODE 377
QY 108 -----KTRAEVKPEGDTLVNRIELKGIDPKEDGNILGHKLEYNNSHNVI 153
DB 378 STWEPDVRWDYADKTR--EPE---LNSAEFDGLDYQDAFEVLAERPE----- 421
QY 154 MADKQKNG-IKYNFKIRHNIEDGSVQLADHYQQNTPI-----GDGPVLLPDN 199
DB 422 ---RQGRQRRVNYRLR-----DWGVSQRVYGCPIVYICPTCGAVPVEDQLPVILPEN 474
QY 200 -HYLSTQSALSADPNEKR 216
DB 475 VAFSGTGSPIKTDPEWRK 492

RESULT 6
D153_HAEIN
ID D153_HAEIN STANDARD; PRT; 793 AA.
AC O32629;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) Outer membrane protein D15).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAK 12085;
RX MEDLINE=97427952; PubMed=9284140;
RA Loomore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease.";
RL Infect Immun 65:3701-3707(1997).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: Belongs to the surface antigen D15 family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U60834; AAB61977.1; -.
DR InterPro; IPR000184; Bac_surfAg_D15.
DR Pfam; PF01103; Bac surface Ag; 1.
KW Antigen; Outer membrane; Signal.
```

```
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 793 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 793 AA; 87511 MW; 51BFDB2036801A14 CRC64;

Query Match
Best Local Similarity 22.8%; Pred. No. 11;
Matches 50; Conservative 27; Mismatches 79; Indels 63; Gaps 11;

QY 65 LSVGVQCFSRYPDHMKQHDFF-----FKSAMPEGYVOE-----RTI 99
DB 427 IGYGTESGISYQTSIKQDNFLGTGAAGVSTAGTKNDYGTSVNLGYTEPYFTKDGVSIGCNI 486
QY 100 FFKDDGNYKTRAEVKPEGDTLVNRIELKGIDPKEDGNI---LGH-----KLEYNNS 148
DB 487 FFENYDNSKSDTSSNYKRTTYGSNVTL-GFPVNNNSYVVLGHTYKNISFALEYN--- 542
QY 149 HNVYIMADKQK-NGIKVNFKIRHNIEDGSVQLADHYQQ-----NTPIGDGPVLL 196
DB 543 RNLYIQSMKFKNGIKTN-----DFDFSFGWYNSLNRYFTPKGVKASLG-GRVTI 593
QY 197 P--DNHYLSTQSALSADPNEKRDMVLVGVFVTAAGITLG 233
DB 594 PGSDNKYYKLSADVQGFYPLDRDRHWVWSAKASAGYANG 632

RESULT 7
GUN_BACS6
ID GUN_BACS6 STANDARD; PRT; 941 AA.
AC P19424;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Alkaline cellulase).
OS Bacillus sp. (strain KSM-635).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1415;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91037937; PubMed=2230718;
RA Ozaki K., Shikata S., Kawai S., Ito S., Okamoto K.;
RT "Molecular cloning and nucleotide sequence of a gene for alkaline
RT cellulase from Bacillus sp. KSM-635.";
RL J. Gen. Microbiol. 136:1327-1334(1990).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M27420; AAA22304.1; -.
DR PIR; S29043; S29043.
DR PDB; 1G01; 31-DEC-02.
DR PDB; 1G0C; 31-DEC-02.
DR InterPro; IPR005086; CBM_17_28.
DR InterPro; IPR008979; Gal_bind like.
DR InterPro; IPR001547; Glyco_hydro_5.
DR InterPro; IPR001119; SLH.
DR Pfam; PF03424; CBM 17 28; 2.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00395; SLH; 3.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR PROSITE; PS01072; SLH DOMAIN; 2.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat;
KW 3D-structure.
```





CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; A5008725; AAL19599.1; -.  
DR StyGene; SG????; leuS.  
DR HAMAP; MF 00049; -; 1.  
DR InterPro; IPR002302; Leu-TRNAsyntla.  
DR InterPro; IPR002300; tRNA-synt\_1a.  
DR InterPro; IPR001412; tRNA-synt\_1.  
DR InterPro; IPR009008; ValRS\_fiers\_edit.  
DR Pfam; PF00133; tRNA-synt\_1; 1.  
DR PRINTS; PR00985; TRNASYNTHLEU.  
DR TIGRFAMs; TIGR00396; leuS\_bact; 1.  
DR PROSITE; PS00178; AA-TRNA-LIGASE I; 1.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KW Complete proteome.  
FT SITE 42 52 "HIGH" REGION.  
FT SITE 619 623 "KMSKS" REGION.  
FT BINDING 622 622 ATP (BY SIMILARITY).  
FT SEQUENCE 860 AA; 96985 MW; D5003584DFECCAB6 CRC64;  
SQ  
Query Match 6.9%; Score 87.5; DB 1; Length 860;  
Best Local Similarity 23.3%; Pred.No.15;  
Matches 42; Conservative 21; Mismatches 78; Indels 39; Gaps 7;  
QY 50 TTGKLPVPTLTTLTSLYGVQCFSPYDHMKQHDFFKSAMPEGYQVETIFFKDDGNYKT 109  
DB 314 TGEIEPV-WAANFVLMYEGTGAVMAVPGH-DQRD-YEFASKYGLTIKPVILAADGSEPD 370  
QY 110 RAEVKPEGDTLVNRIELKIDFKEDCNILGHKLEYNHNVYIMADKQNGIKVNFKR 169  
DB 371 SEQALTEKGVLFNSGFEFDGLAEAFNAIADKL-----AEKGVGERKVNRLR 418  
QY 170 H-----NIEDGSVLADHYQQNTPIGDGVLPLPDNHYL-STQSALSKDP 212  
DB 419 DWGVSQRVYGAPIPWTTLEDGTV-----LPTPEDQLPVILPDVMDGITSPIKADP 471  
RESULT 10  
ID DPOE YEAST STANDARD; PRT; 2222 AA.  
AC P21951;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE DNA polymerase epsilon, catalytic subunit A (EC 2.7.7.7) (DNA  
DE polymerase II subunit A).  
GN POL2 OR DUN2 OR YNL262W OR N0825.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1214-1221.  
RX MEDLINE=90381771; PubMed=2169349;  
RA Morrison A., Araki H., Clark A.B., Hamatake R.K., Sugino A.;  
RT "A third essential DNA polymerase in S. cerevisiae";  
RL Cell 62:1143-1151(1990).  
RN [2]  
RP SEQUENCE OF 1-2221 FROM N.A.  
RC STRAIN=S288c / FY1679;  
RX MEDLINE=96310631; PubMed=8740425;  
RA Sen-Gupta M., Lyck R., Fleig U., Nidenthal R.K., Hegemann J.H.;  
RT "The sequence of a 24,152 bp segment from the left arm of chromosome  
RT XIV from Saccharomyces cerevisiae between the BNI1 and the POL2

RT Genes.";  
RL Yeast 12:505-514 (1996).  
RN [3]  
RP TEMPERATURE SENSITIVE MUTANTS.  
RX MEDLINE=92164663; PubMed=1537345;  
RA Araki H., Kopp P.A., Johnson A.L., Johnston L.H., Morrison A.,  
RA Sugino A.;  
RT "DNA polymerase II, the probable homolog of mammalian DNA polymerase  
RT epsilon, replicates chromosomal DNA in the yeast Saccharomyces  
RT cerevisiae";  
RL EMBL J. 11:733-740 (1992).  
CC -!- FUNCTION: DNA POLYMERASE II PARTICIPATES IN CHROMOSOMAL DNA  
CC REPLICATION.  
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
CC + [DNA] (N).  
CC -!- SUBUNIT: CONSISTS OF FIVE SUBUNITS (200 kDa, 80 kDa, 34 kDa, 30  
CC kDa, AND 29 kDa).  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- DOMAIN: THE DNA POLYMERASE ACTIVITY DOMAIN RESIDES IN THE  
CC N-TERMINAL HALF OF THE PROTEIN, WHILE THE C-TERMINUS IS NECESSARY  
CC FOR COMPLEXING SUBUNITS B AND C.  
CC -!- MISCELLANEOUS: In eukaryotes there are five DNA polymerases:  
CC alpha, beta, gamma, delta, and epsilon which are responsible for  
CC different reactions of DNA synthesis.  
CC -!- SIMILARITY: Belongs to the DNA polymerase type-B family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M60416; AAA88711.1; -.  
DR EMBL; X92494; CAA63235.1; -.  
DR EMBL; Z71538; CAA96169.1; -.  
DR PIR; A36028; A36028.  
DR GeneOnline; 143268; -.  
DR SGD; S0005206; POL2.  
DR GO; GO:0000731; P:DNA repair synthesis; IMP.  
DR InterPro; IPR006172; DNA\_pol\_B.  
DR InterPro; IPR006134; DNA\_pol\_B\_dom.  
DR InterPro; IPR006133; DNA\_pol\_B\_exo.  
DR Pfam; PF00136; DNA\_pol\_B\_1.  
DR Pfam; PF03104; DNA\_pol\_B\_exo; 1.  
DR SMART; SM00486; POLB; 1.  
KW TRANSFERASE; DNA-directed DNA polymerase; DNA replication;  
KW DNA-binding; Zinc-finger; Nuclear protein.  
FT ZN FING 2108 2181 POTENTIAL.  
FT VARIANT 644 644 M -> I (IN POL2-9 TS MUTANT).  
FT VARIANT 710 710 P -> S (IN POL2-18 TS MUTANT).  
SQ SEQUENCE 2222 AA; 255669 MW; CBCDDE2AB147D65B CRC64;  
Query Match 6.9%; Score 87.5; DB 1; Length 2222;  
Best Local Similarity 28.2%; Pred.No.46;  
Matches 37; Conservative 14; Mismatches 49; Indels 31; Gaps 7;  
QY 54 LPVPWP-TLVTTLSYGVQCFSPYDHM-----KQHDFFKSAMPEGYV-----QERTI 99  
DB 883 LPKSPETVFFFTLENGKKLYLSPCSMLNRYVHQFTNHQYQELKDPNLNVIYTHSENTI 942  
QY 100 FFKDDGNYKTR--AEVKFEGDTLVNR-----IELKGDIDFKEDGNILGHKLEYN 147  
DB 943 FFEVDGPKYKAMILPSKKEGKIKRYAVFNEDGSLAELKGFELKRGEL---QLIKNFQ 999  
QY 148 S--HNVYIMAD 156  
DB 1000 SDIFKVFLEGD 1010  
RESULT 11



```
Db 473 FFGIVPVLMDKGSV-----VEGSNVSGALCIS-----QAWPGMARTI--- 510
QY 69 VQCFSRYPDHMKQHDFFKSPGPGYQERTIFFKDGNYKTRA---EVKPEGDTLVNRIE 125
Db 511 -----YGDHQRFDVAYFKRYP-GY-----YFTGDGAYTEGGYQITGRMDVI----- 553
QY 126 LKGIDPFKDGNIHGKHL 142
Db 554 -----NISGHRLL 560

RESULT 12
CP51_CANGA STANDARD; PRT; 533 AA.
ID CP51_CANGA
AC P50859; Q02312;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 51 (EC 1.14.13.70) (CYPLI) (P450-LIAl) (Sterol 14-
alpha-demethylase) (lanosterol 14-alpha demethylase) (P450-14DM).
GN ERG11 OR CYP51.
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2001-L5;
RX MEDLINE=96161286; PubMed=8593007;
RA Geber A., Hitchcock C.A., Swartz J.E., Pullen F.S., Marsden K.E.,
RA Kwon-Chung K.J., Bennett J.E.;
RT "Deletion of the Candida glabrata ERG3 and ERG11 genes: effect on cell
RT viability, cell growth, sterol composition, and antifungal
RT susceptibility.";
RL Antimicrob. Agents Chemother. 39:2708-2717(1995).
RN [2]
RP SEQUENCE OF 60-473 FROM N.A.
RC STRAIN=ATCC 2001;
RX MEDLINE=95081364; PubMed=7969540;
RA Burgener-Kairuz P., Zuber J.P., Jaunin P., Buchman T.G., Bille J.,
RA Rosbier M.;
RT "Rapid detection and identification of Candida albicans and
RT Torulopsis (Candida) glabrata in clinical specimens by
RT species-specific nested PCR amplification of a cytochrome P-450
RT lanosterol-alpha-demethylase (LIAl) gene fragment.";
RL J. Clin. Microbiol. 32:1902-1907(1994).
CC -!- FUNCTION: Catalyzes C14-demethylation of lanosterol which is
CC critical for ergosterol biosynthesis. It transforms lanosterol
CC into 4,4'-dimethyl cholesterol-8,14,24-triene-3-beta-ol (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: Obtusifoliol + 3 O(2) + 3 NADPH = 4-alpha-
CC methyl-5-alpha-ergosta-8,14,24(28)-trien-3-beta-ol + formate + 3
CC NADP(4+) + 3 H(2)O.
CC -!- PATHWAY: Ergosterol biosynthesis.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; L40389; AAB02329.1; -.
CC EMBL; S75389; AAB32679.1; -.
CC InterPro; IPR001128; Cytochrome_P450.
CC Pfam; PF00067; P450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Electron transport; Oxidoreductase; Monooxygenase; Membrane; Heme;
KW Sterol biosynthesis; NADP.
FT METAL 472 472 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
```

```
FT CONFLICT 64 64 I -> M (IN REF. 2).
FT CONFLICT 473 473 I -> T (IN REF. 2).
SQ SEQUENCE 533 AA; 61305 MW; A0506C17507E6EF7 CRC64;

Query Match 6.8%; Score 86.5; DB 1; Length 533;
Best Local Similarity 21.4%; Pred. No. 10;
Matches 44; Conservative 32; Mismatches 81; Indels 49; Gaps 8;

QY 25 GHKFSVS---GEGEGDATYKGLTKFKICTGKLPVWPMTLVTTLSYGVQCFSRYPDH--M 79
Db 109 GHEFIENAKLADVSAEAAYSHL-----TTPVFGKGVYDCPNRLM 149
QY 80 KOHDFFSAM-PEGYV-----QERTIFFKDGNYKTRAEVKPEGDTLVNRIELKGIDF 131
Db 150 EQKKFVKGALTKEAFVRYVPLIAEIIKYFRNSKFNENNSGIVDMVMSQPEM--TIF 207
QY 132 KEDGNILGHKLEYNVSHNVYIMADKQNGIKVFKIRHNIEDSGVQLADHYQNTPIGD 191
Db 208 TARSRLLGKEMRDKLDTPFAYLYSLDKGFTPINF-VFPNLPLEHYKRDHQAQAIS--- 263
QY 192 GPVLLPDNHYLSTQSALSQDPNEKRD 217
Db 264 -----GTYSMLIKERKND 278

RESULT 13
SYL_XYLFT STANDARD; PRT; 879 AA.
ID SYL_XYLFT
AC Q87C65;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR PDI230.
OS Xylella fastidiosa (strain Temeculal / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=183190;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22421331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorty H., Tsai S.M.,
RA Carver H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA Coutinho L.B., Kimura E.F., Ferro E.S., Harakava R., Kuramae E.E.,
RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,
RA da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sasaki F.T., Sena J.A.D.,
RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,
RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA Kitajima J.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of Xylella
RT fastidiosa";
RL J. Bacteriol. 185:1018-1026(2003).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AF012557; AAO29080.1; ALT_INIT.
CC HAMAP; MF_00049; -; 1.
DR
```

```
DR InterPro; IPR002302; Leu-TRNAsyntla.
DR InterPro; IPR002300; tRNA-synt 1a.
DR InterPro; IPR001412; tRNA-synt 1.
DR InterPro; IPR009008; ValRS_fiers_edit.
DR Pfam; PF00133; tRNA-synt 1; 1.
DR PRINTS; PRO0985; TRNASYNTHLSU.
DR TIGRFAMS; TIGR00396; leuS bact; 1.
DR PROSITE; PS00178; AA TRNA LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 45 55 "HIGH" REGION.
FT SITE 637 641 "KMSKS" REGION.
FT BINDING 640 640 ATP (BY SIMILARITY).
FT SEQUENCE 879 AA; 99823 MW; 4C2E01B8FDC497E CRC64;
SQ
Query Match 6.8%; Score 86.5; DB 1; Length 879;
Best Local Similarity 22.2%; Pred. No. 18;
Matches 44; Conservative 28; Mismatches 69; Indels 57; Gaps 10;
QY 50 TTGKLPVWPTLVTLSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNY-- 107
DB 321 TNEQLPV-WVANFVLMAYGTGAVMVGPHDQDEF--ANKYGLPIRQVIALKEPKNQDE 377
QY 108 -----KTRAEVKFEGDTLVNRIELKIDFKEDGNILGHKLEYNNSHVYI 153
DB 378 SIWEPDVRWDYADKTR---EFE--LINSAPFDGLDYQGAPEVLAERFE----- 421
QY 154 MADKQKNG-IKVNFKIRHNIEDGSVOLADHYQONTPI-----GDGPVLLPON 199
DB 422 ---RQGRQRRVYRLR----DWGVSQRQYWGCFPIEVIYCTGCAVPVPENQLPVILPEN 474
QY 200 -HYLSTQSALS KDPNEKR 216
DB 475 VAPSGTGSPKIKDPFEWRK 492
RESULT 14
VIT4_CAREL STANDARD; PRT; 1603 AA.
AC FL8947; Q9BPP3;
DT 01-NOV-1990 (Rel. 16, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vitellogenin 4 precursor.
VIT-4 OR F59D8.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-282 FROM N.A.
RA Blumenthal T.; Spieth J.; Zucker E.;
RL Submitted (NOV-1986) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-71 FROM N.A.
RX MEDLINE=85269643; PubMed=4022780;
RA Spieth J.; Denison K.; Kirtland S.; Cane J.; Blumenthal T.;
RT "The C. elegans vitellogenin genes: short sequence repeats in the
RT promoter regions and homology to the vertebrate genes."
RL Nucleic Acids Res. 13:5283-5295(1985).
CC -1- FUNCTION: Precursor of the egg-yolk proteins that are sources of
CC nutrients during embryonic development (Potential).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Synthesized in Caenorhabditis only by 32 cells
CC building the intestine of adult hermaphroditic individuals; they
CC are cotranslationally secreted into the body cavity and
CC subsequently taken up by the gonad.
CC -1- SIMILARITY: Contains 1 VWFD domain.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; AC024137; AAK09074.1; -.
DR EMBL; M11498; AAZ28163.1; -.
DR EMBL; X02754; CAA26531.1; -.
DR PIR; A43084; A43084.
DR WormPep; F59D8.2; CE26817.
DR InterPro; IPR001747; Lipid_transprt_N.
DR InterPro; IPR001846; VWFD.
DR Pfam; PF01347; Vitellogenin_N; 1.
DR SMART; SM00638; LPD_N; 1.
DR SMART; SM00216; VMD; 1.
KW Storage protein; Multigene family; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 1603 VWFD.
FT DOMAIN 1308 1455
FT CONFLICT 30 30 Y -> V (IN REF. 3).
FT CONFLICT 169 169 L -> V (IN REF. 2).
FT CONFLICT 183 187 EVAYT -> RSLRH (IN REF. 2).
FT CONFLICT 275 275 T -> S (IN REF. 2).
SQ SEQUENCE 1603 AA; 186307 MW; E303170325BC99BB CRC64;
Query Match 6.8%; Score 86.5; DB 1; Length 1603;
Best Local Similarity 23.4%; Pred. No. 37;
Matches 52; Conservative 32; Mismatches 69; Indels 69; Gaps 12;
QY 1 MYSKGEELFTGVVPIILVELDGVNKGKPSVGEDEGATYGLTLKFICTTGKLPVWPWT 60
DB 162 MESDKDLSFFNVHEKTMGDCV---AYTIIVEG-GKTIYTSVNFDKCITR-----PE 211
QY 61 LVTTLISYGVQCFSRYPDHMKQHDFFKSAMPEG-YVQERTIF---FKDDG----- 105
DB 212 TAYGLRFGSEC-----KECKEGQFVQPTVTVTFKNEKLEQSEVNSIYT 257
QY 106 -----NYKTRAVKFEGETLVNRIELKIDFKEDGNILGHKLEYNNSHVIMAD 156
DB 258 LNVNGQEVVKSETRAKTVFEESKINR-EIK-----KVGPKKEIYVSMENKLEIQ 308
QY 157 KQKNG-----IKVNFKIRHNIEDGSVOLADHYQONTP 188
DB 309 FYQGDKAENVFPKALIEQKV-EQLEEIFRQIQEH-EQNTTP 348
RESULT 15
TRNA_CAMJE STANDARD; PRT; 357 AA.
ID TRNA_CAMJE
AC Q9PP92;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE tRNA (Uracil-5)-methyltransferase (EC 2.1.1.35) (tRNA(M-5-U54)-
DE methyltransferase) (RUMT).
GN TRNA OR CJ0831C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J.; Wren B.W.; Mungall K.; Ketley J.M.; Churcher C.;
RA Basham D.; Chillingworth T.; Davies R.M.; Feltwell T.; Holtroyd S.;
RA Jagels K.; Karlyshev A.V.; Moule S.; Pallen M.J.; Penn C.W.;
RA Quail M.A.; Rajandream M.A.; Rutherford K.M.; van Vliet A.H.M.;
RA Whitehead S.; Barrell B.G.;
```





Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	1240	97.3	238	2	Q8GHE2	Q8ghe2 azotobacter
2	1237	97.1	238	5	Q93125	Q93125 aequorea vi
3	1235	96.9	238	2	Q8GHE4	Q8ghe4 azomonas ag
4	1234	96.9	238	2	Q8GHE3	Q8ghe3 azotobacter
5	1202	94.3	238	5	Q17105	Q17105 aequorea vi
6	1187	93.2	238	5	Q17106	Q17106 aequorea vi
7	1082	84.9	238	5	Q8WTC6	Q8wtc6 aequorea ma
8	1078	84.6	238	5	Q8WP95	Q8wp95 aequorea ma
9	1074	84.3	238	5	Q8WTC4	Q8wtc4 aequorea ma
10	1072	84.1	238	5	Q8WTD0	Q8wtd0 aequorea ma
11	1071	84.1	238	5	Q8WTC8	Q8wtc8 aequorea ma
12	1071	84.1	238	5	Q8WTC9	Q8wtc9 aequorea ma
13	1069	83.9	238	5	Q8WTC7	Q8wtc7 aequorea ma
14	1067	83.8	238	5	Q8WTC5	Q8wtc5 aequorea ma
15	252.5	19.8	225	5	Q95UA7	Q95ua7 montastraea
16	252.5	19.8	225	5	Q720W5	Q720w5 montastraea

```

Db 61 VTTFSYGVQCFSRYPDHMKRHHDFKSAPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120
QY 122 NRIELKGIQDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
Db 121 NRIELKGIQDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180
QY 182 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVFTAAAGITLGMDELYK 239
Db 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVFTAAAGITLGMDELYK 238

RESULT 2
Q93125 PRELIMINARY; PRT; 238 AA.
AC Q93125;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein mutant 3.
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96305137; PubMed=8707053;
RA Cormack B.P., Valdivia R.H., Falkow S.;
RT "FACS-optimized mutants of the green fluorescent protein (GFP).";
RL Gene 173:33-38(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Cormack B.P., Bertram G., Egerton M., Gow N.A.R., Falkow S.,
RA Brown A.J.P.;
RT "Yeast Enhanced Green Fluorescent Protein (yEGFP): a reporter of gene
RT expression in Candida albicans.";
RL Microbiology 0:0-0(1996).
DR EMBL; U73501; AAB18957.1; -.
DR HSSP; P42212; 1BFP.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26840 MW; A28622809A9DEA60 CRC64;

Query Match 97.1%; Score 1237; DB 5; Length 238;
Best Local Similarity 97.1%; Pred. No. 8,7e-96;
Matches 231; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 61
Db 1 MSXGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 60
QY 62 VTTLSYGVQCFSRYPDHMKRHHDFKSAPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
Db 61 VTTFSYGVQCFSRYPDHMKRHHDFKSAPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120
QY 122 NRIELKGIQDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
Db 121 NRIELKGIQDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180
QY 182 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVFTAAAGITLGMDELYK 239
Db 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVFTAAAGITLGMDELYK 238

RESULT 3
Q8GHE4 PRELIMINARY; PRT; 238 AA.
AC Q8GHE4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)

```

```

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescence protein.
GN 375GFP.
OS Azomonas agilis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azomonas.
OX NCBI_TaxID=116849;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
RT Azomonas and Azotobacter.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324405; AAN86137.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26902 MW; 15FE9B9C5B4F6B89 CRC64;

Query Match 96.9%; Score 1235; DB 2; Length 238;
Best Local Similarity 97.1%; Pred. No. 1.3e-95;
Matches 231; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 61
Db 1 MSXGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 60
QY 62 VTTLSYGVQCFSRYPDHMKRHHDFKSAPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 121
Db 61 VTTFSYGVQCFSRYPDHMKRHHDFKSAPEGYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120
QY 122 NRIELKGIQDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 181
Db 121 NRIELKGIQDFKEDGNILGHKLEYNYNHNVYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180
QY 182 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVFTAAAGITLGMDELYK 239
Db 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDMHVLGVFTAAAGITLGMDELYK 238

RESULT 4
Q8GHE3 PRELIMINARY; PRT; 238 AA.
AC Q8GHE3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescence protein.
GN 85GFP.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RA Koranyi P., Berenyi M., Burg K.;
RT "Occurrence of green fluorescence protein in diazotrophic bacteria
RT Azomonas and Azotobacter.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF324406; AAN86138.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26887 MW; E0E1616BD2AF6188 CRC64;

Query Match 96.9%; Score 1234; DB 2; Length 238;

```



## RESULT 6



RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR ENBL1: AFA333912.1; P:energy; GO: GO:0006091; P:energy pathways; IEA.  
DR InterPro: IPR009017; GFP like.  
DR InterPro: IPR000786; Green\_fl\_protein.  
DR Pfam: PF01353; GFP; 1.  
DR PRINTS: PR01229; GFP; 1.  
DR P-codon: PD013756; Green fl protein; 1.  
SO SEQUENCE 238 AA: 26997 MW: 5780A1921773CB84D CRC64:

Query Match	Score 1072;	DB 5;	Length 238;
Best Local Similarity	81.1%;	Pred. No. 6e-82;	
Matches 193; Conservative	21;	Mismatches 24;	Indels 0; Gaps 0;

Qy	2	VS	KG	EE	LT	GV	VP	IL	VE	LD	GD	VN	GH	KF	VS	GE	GD	AT	VG	KL	TL	KFI	CT	TG	KL	VP	VP	WT	61
Dd	1	MS	KG	EE <td>LT</td> <td>GV <td>VP <td>IL</td> <td>VE <td>LD <td>GD <td>VN <td>GH <td>KF <td>VS <td>GE <td>GD <td>AT</td> <td>VG <td>KL <td>TL</td> <td>KFI</td> <td>CT</td> <td>TG <td>KL</td> <td>VP <td>VP <td>WT</td> <td>60</td> </td></td></td></td></td></td></td></td></td></td></td></td></td></td></td></td>	LT	GV <td>VP <td>IL</td> <td>VE <td>LD <td>GD <td>VN <td>GH <td>KF <td>VS <td>GE <td>GD <td>AT</td> <td>VG <td>KL <td>TL</td> <td>KFI</td> <td>CT</td> <td>TG <td>KL</td> <td>VP <td>VP <td>WT</td> <td>60</td> </td></td></td></td></td></td></td></td></td></td></td></td></td></td></td>	VP <td>IL</td> <td>VE <td>LD <td>GD <td>VN <td>GH <td>KF <td>VS <td>GE <td>GD <td>AT</td> <td>VG <td>KL <td>TL</td> <td>KFI</td> <td>CT</td> <td>TG <td>KL</td> <td>VP <td>VP <td>WT</td> <td>60</td> </td></td></td></td></td></td></td></td></td></td></td></td></td></td>	IL	VE <td>LD <td>GD <td>VN <td>GH <td>KF <td>VS <td>GE <td>GD <td>AT</td> <td>VG <td>KL <td>TL</td> <td>KFI</td> <td>CT</td> <td>TG <td>KL</td> <td>VP <td>VP <td>WT</td> <td>60</td> </td></td></td></td></td></td></td></td></td></td></td></td></td>	LD <td>GD <td>VN <td>GH <td>KF <td>VS <td>GE <td>GD <td>AT</td> <td>VG <td>KL <td>TL</td> <td>KFI</td> <td>CT</td> <td>TG <td>KL</td> <td>VP <td>VP <td>WT</td> <td>60</td> </td></td></td></td></td></td></td></td></td></td></td></td>	GD <td>VN <td>GH <td>KF <td>VS <td>GE <td>GD <td>AT</td> <td>VG <td>KL <td>TL</td> <td>KFI</td> <td>CT</td> <td>TG <td>KL</td> <td>VP <td>VP <td>WT</td> <td>60</td> </td></td></td></td></td></td></td></td></td></td></td>	VN <td>GH <td>KF <td>VS <td>GE <td>GD <td>AT</td> <td>VG <td>KL <td>TL</td> <td>KFI</td> <td>CT</td> <td>TG <td>KL</td> <td>VP <td>VP <td>WT</td> <td>60</td> </td></td></td></td></td></td></td></td></td></td>	GH <td>KF <td>VS <td>GE <td>GD <td>AT</td> <td>VG <td>KL <td>TL</td> <td>KFI</td> <td>CT</td> <td>TG <td>KL</td> <td>VP <td>VP <td>WT</td> <td>60</td> </td></td></td></td></td></td></td></td></td>	KF <td>VS <td>GE <td>GD <td>AT</td> <td>VG <td>KL <td>TL</td> <td>KFI</td> <td>CT</td> <td>TG <td>KL</td> <td>VP <td>VP <td>WT</td> <td>60</td> </td></td></td></td></td></td></td></td>	VS <td>GE <td>GD <td>AT</td> <td>VG <td>KL <td>TL</td> <td>KFI</td> <td>CT</td> <td>TG <td>KL</td> <td>VP <td>VP <td>WT</td> <td>60</td> </td></td></td></td></td></td></td>	GE <td>GD <td>AT</td> <td>VG <td>KL <td>TL</td> <td>KFI</td> <td>CT</td> <td>TG <td>KL</td> <td>VP <td>VP <td>WT</td> <td>60</td> </td></td></td></td></td></td>	GD <td>AT</td> <td>VG <td>KL <td>TL</td> <td>KFI</td> <td>CT</td> <td>TG <td>KL</td> <td>VP <td>VP <td>WT</td> <td>60</td> </td></td></td></td></td>	AT	VG <td>KL <td>TL</td> <td>KFI</td> <td>CT</td> <td>TG <td>KL</td> <td>VP <td>VP <td>WT</td> <td>60</td> </td></td></td></td>	KL <td>TL</td> <td>KFI</td> <td>CT</td> <td>TG <td>KL</td> <td>VP <td>VP <td>WT</td> <td>60</td> </td></td></td>	TL	KFI	CT	TG <td>KL</td> <td>VP <td>VP <td>WT</td> <td>60</td> </td></td>	KL	VP <td>VP <td>WT</td> <td>60</td> </td>	VP <td>WT</td> <td>60</td>	WT	60

[illegible][illegible]

Qy  
182 HYQONTPTGDPVLLPDNNHLSQTQSALSKDPNEKRDDHMLVGVFTAAAGTTLGMDLEYK 239  
181 HYQTNVPLGDPVLPINHYLSLQTAISKORNETRDHMLVLEFPSCAGTHGMDLEYK 238  
Db

RESULT 11	
Q8WTC8	
ID Q8WTC8	DDPYTMINADY.
	DPT.
	230 AA

AC	Q8WTC8;
DT	(TREMBlrel. 20, Created)
DT	(TREMBlrel. 20, Last sequence update)
DT	(TREMBlrel. 20, Last annotation update)
DT	(TREMBlrel. 25, Last annotation update)

DE green fluorescent protein.  
GN GFP.  
OS Aequorea macrodactyla.  
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae.

CC Aequoreidae; Aequorea.  
 OX NCBI\_TaxID=147615;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP

RC  
RA  
RT  
RT  
RT  
RT

SIRALINEOPHILUS;  
Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,  
Li S.J., Xia N.S.;  
"Colorful mutants of green fluorescent protein from Aequorea"

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted January 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

InterPro; IPR000786; Green\_fl\_protein.  
Pfam; PF01353; GPP; 1.  
PRINTS; PR01229; GFLUORESCENT.

```

SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;
Query Match 84.1%; Score 1071; DB 5; Length 238;

```

	Matches	193;	Conservative	21;	Mismatches	24;	Indels	0;	Gaps	0;
Qy	2	VSKGEELFTGV	PILVELDGDV	NHKKFVS	GEGEDATY	GKLT	TKFKIT	CTTGKLP	VPWP	PTL 61

Db	1	MSKGEELFTGIVPVLIELDGDVGHGKPSVRGEGGDADYGLKEIKFICTTGKLPVWP	60
Qy	62	VTTLSYGVQCFSSYPDHMKQHDPEKSNMPEGYQERTTIEFKDGNKYKRAEVKESGD	121

Db 61 VTTLGYGICFARYPEHMKMNDFFKSAMPEGY<sup>1</sup>QERTIFFQDDGKYKTRGEVAFEGDTLV 120

Qy	1 2 2	NRIELKIGDIFKEDGNILGHKLEBYNYSNHNVTYIMADKQKNGIKVNFKFIHNIEDGVSQVLAD	181
Db	1 2 1	NRIELKGMDFREDGNILGHKLEYNFNSHNVTYIMPDKANGLKGFNFKFIHNIIEGGVSQVLAD	180

Qy	182	HYQONTPIGDGEVLLPDNHYLSTQSALSKDPNEKRDHMLVGFVTAAGITLGNDELYK	239
		:     :     :     :     :     :     :     :     :     :	
Dd	181	HYQTNVPLGDGEVLLPDNHYLSYQTALSKDRNETRDHMLVEFFFSACGHTGNDELYK	238
		:     :     :     :     :     :     :     :     :	

RESULT 12  
Q8WTC9

AC	Q8WTC9:	(TREMBLrel. 20, Created)
DT	01-MAR-2002	(TREMBLrel. 20, Last sequence update)
DT	01-MAR-2002	

DE Green fluorescent protein.  
GN GFP.  
OS Aequorea macrodactyla.  
C Aequorea macrodactyla.

OC Aequoreidae; Aequorea.  
 NCBI\_TaxID=147615;  
 [1]  
 EN DD SEQUENCE FROM N A

RC STRAIN=GFPxM162;  
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,  
RA Li S.J., Xia N.S.;  
RA "Colorful mutants" of green fluorescent protein from *Aequorea*  
BT "Colorful mutants" of green fluorescent protein from *Aequorea*

RT macrodactyla.";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF435428; AAL33913.1; -;  
DR GO: GO:0006091; p-energy pathways; IEA.

DK InterPro; IPR009017; GFP-like.  
DR InterPro; IPR000786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFP; 1.  
DR PRINTS; PR01229; GFP; 1.

DR	Protein	Score	DB	Length
1	Protein, 1.1	84.1%	Score 1071;	Length 238;
2	Protein, 2.2	84.1%	Score 1071;	Length 238;
3	Protein, 3.3	84.1%	Score 1071;	Length 238;
4	Protein, 4.4	84.1%	Score 1071;	Length 238;
5	Protein, 5.5	84.1%	Score 1071;	Length 238;
6	Protein, 6.6	84.1%	Score 1071;	Length 238;
7	Protein, 7.7	84.1%	Score 1071;	Length 238;
8	Protein, 8.8	84.1%	Score 1071;	Length 238;
9	Protein, 9.9	84.1%	Score 1071;	Length 238;
10	Protein, 10.10	84.1%	Score 1071;	Length 238;
11	Protein, 11.11	84.1%	Score 1071;	Length 238;
12	Protein, 12.12	84.1%	Score 1071;	Length 238;
13	Protein, 13.13	84.1%	Score 1071;	Length 238;
14	Protein, 14.14	84.1%	Score 1071;	Length 238;
15	Protein, 15.15	84.1%	Score 1071;	Length 238;
16	Protein, 16.16	84.1%	Score 1071;	Length 238;
17	Protein, 17.17	84.1%	Score 1071;	Length 238;
18	Protein, 18.18	84.1%	Score 1071;	Length 238;
19	Protein, 19.19	84.1%	Score 1071;	Length 238;
20	Protein, 20.20	84.1%	Score 1071;	Length 238;
21	Protein, 21.21	84.1%	Score 1071;	Length 238;
22	Protein, 22.22	84.1%	Score 1071;	Length 238;
23	Protein, 23.23	84.1%	Score 1071;	Length 238;
24	Protein, 24.24	84.1%	Score 1071;	Length 238;
25	Protein, 25.25	84.1%	Score 1071;	Length 238;
26	Protein, 26.26	84.1%	Score 1071;	Length 238;
27	Protein, 27.27	84.1%	Score 1071;	Length 238;
28	Protein, 28.28	84.1%	Score 1071;	Length 238;
29	Protein, 29.29	84.1%	Score 1071;	Length 238;
30	Protein, 30.30	84.1%	Score 1071;	Length 238;
31	Protein, 31.31	84.1%	Score 1071;	Length 238;
32	Protein, 32.32	84.1%	Score 1071;	Length 238;
33	Protein, 33.33	84.1%	Score 1071;	Length 238;
34	Protein, 34.34	84.1%	Score 1071;	Length 238;
35	Protein, 35.35	84.1%	Score 1071;	Length 238;
36	Protein, 36.36	84.1%	Score 1071;	Length 238;
37	Protein, 37.37	84.1%	Score 1071;	Length 238;
38	Protein, 38.38	84.1%	Score 1071;	Length 238;
39	Protein, 39.39	84.1%	Score 1071;	Length 238;
40	Protein, 40.40	84.1%	Score 1071;	Length 238;
41	Protein, 41.41	84.1%	Score 1071;	Length 238;
42	Protein, 42.42	84.1%	Score 1071;	Length 238;
43	Protein, 43.43	84.1%	Score 1071;	Length 238;
44	Protein, 44.44	84.1%	Score 1071;	Length 238;
45	Protein, 45.45	84.1%	Score 1071;	Length 238;
46	Protein, 46.46	84.1%	Score 1071;	Length 238;
47	Protein, 47.47	84.1%	Score 1071;	Length 238;
48	Protein, 48.48	84.1%	Score 1071;	Length 238;
49	Protein, 49.49	84.1%	Score 1071;	Length 238;
50	Protein, 50.50	84.1%	Score 1071;	Length 238;
51	Protein, 51.51	84.1%	Score 1071;	Length 238;
52	Protein, 52.52	84.1%	Score 1071;	Length 238;
53	Protein, 53.53	84.1%	Score 1071;	Length 238;
54	Protein, 54.54	84.1%	Score 1071;	Length 238;
55	Protein, 55.55	84.1%	Score 1071;	Length 238;
56	Protein, 56.56	84.1%	Score 1071;	Length 238;
57	Protein, 57.57	84.1%	Score 1071;	Length 238;
58	Protein, 58.58	84.1%	Score 1071;	Length 238;
59	Protein, 59.59	84.1%	Score 1071;	Length 238;
60	Protein, 60.60	84.1%	Score 1071;	Length 238;
61	Protein, 61.61	84.1%	Score 1071;	Length 238;
62	Protein, 62.62	84.1%	Score 1071;	Length 238;
63	Protein, 63.63	84.1%	Score 1071;	Length 238;
64	Protein, 64.64	84.1%	Score 1071;	Length 238;
65	Protein, 65.65	84.1		

QY 2 VSKGEELFTGVVPILVELDGDVNGHKRFSVSGEGEGDATYGKLTCLKTCTTGKLPVPMPTL 61  
Matches 193; Conservative 24; Mismatches 24; Indels 0; Gaps 0;  
Best Local Similarity 0.71; Pos. 1; Neg. 1; Pos. 2; Neg. 2;  
Best Local Similarity 0.71; Pos. 1; Neg. 1; Pos. 2; Neg. 2;

Db	1	MSKGEELFTGIVPVLIIITELDGDVHGHRFSVRGEGEDADYKGLKIKFTCTTGKLPVPWPTL	60
Qy	62	VTTLSYGVQCSFRYPDHMKQHDFFKKSAMPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTLV	121

61	VTTLGIGIOCFARYPEHMKMNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEVKEGDTLV	120
Db		
Qy	122 NRIELKGI DFKBEDNTLGHKL EYNNYNSHNVTIMADKQKNGIKVNFKTRHNEDGSVQLAD	181

Db	121	NRIELKGMDFKEDGNILGHKLEYFNFSHNVI	IMPDKANGLKVNFKTRHNIEGGVQLAD	180
Qy	182	HYQNTPIGDGPVLLPDNHYLSTQSALS	KDPNEKRDMVLVGFVTAAGITGMDELYK	239

Db 181 HYQTNVPLGDGPVLI PINHYLSFQTALSKDRNETRDHMFVLEFFSACGTHGMDELYK 238

Q8WTC7 PRELIMINARY; PRT; 238 AA.

DT 01-MAR-2002 (T-EMBLRel. 20, Last sequence update)  
DT 01-OCT-2003 (T-EMBLRel. 25, Last annotation update)  
DE Green fluorescent protein.  
CN GFP

OS Aequorea macrodactylia.  
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;  
OC Aequoreidae; Aequorea.  
OC NCBI TaxID=147515.  
OX

```

1 MSKGEELFTGVVPIILVELDGDVHGKFSVRGEGGDADYKGLKLEIKFICTTGKLPVPWPPTL 60
62 VTTLSYGVQCSRYPDHMKQHDFFKSAWPEGVQVORTIFFKDDGNVKTAEVKFEGDITLV 121
61 VTTLGYGILCFARYPEHMKMDFFKSAWPEGYIQERTIFFQDDGKYKIRGEVKFEGDITLV 120
122 NRIELKIGDIFKEDGNILGHKLEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLAD 181
121 NRIELKGMDFKEDGNILGHKLEYNFSNHNVIYIMPKANNGLKVNFKIRHNIEGGVQLAD 180
182 HYQONTPTGDDGVLPLDPNHYLSTQALSQDPNEKRDHVMVLGVFTAAAGITLGMDELYK 239
181 HYQTNVPLGDGVLPIPNHYLSYQTAISKDRNETRDHVMVLEFFSACGTHGMDELYK 238

```

RESULT 15  
 Q95UA7 PRELIMINARY; PRT; 225 AA.  
 ID Q95UA7;  
 AC Q95UA7;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Cyan fluorescent protein (fragment).  
 OS Montastraea cavernosa (great star coral).  
 OC Eukaryota; Metazoa; Chnidaria; Anthozoa; Zoantharia; Scleractinia;  
 OC Faviina; Faviidae; Montastraea.  
 NCBI\_TaxID=63558;  
 RN [1]  
 RZ SEQUENCE FROM N.A.  
 RP Falkowski P.G., Sun Y.;  
 RA "Montastraea cavernosa fluorescent protein."  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AV056460; AAL17905.1; -;  
 DR GO; GO:0006091; P.energy pathways; IEA.  
 DR InterPro; IPR009017; GFP\_like.  
 DR InterPro; IPR000786; Green\_fl\_protein.  
 DR Pfam; PF01353; GFP; 1.  
 DR PRINTS; PR01229; GFP\_LUORESCENT.  
 DR ProDom; PD013756; Green\_fl\_protein; 1.  
 FT NON\_TER 225  
 FT SEQUENCE 225 AA; 25775 MW; 52DE2F716D083524 CRC64;

Query Match 19.8%; Score 252.5; DB 5; Length 225;  
 Best Local Similarity 31.1%; Pred. No. 2.9e-13;  
 Matches 64; Conservative 43; Mismatches 82; Indels 17; Gaps 7;

QY 12 VPIILVELDGDVNGHKFSVRGEGDATYKGLTKLF-ICTTGKLPVPWPPTLVTLTSYGVQ 70  
 Db 7 VMKILRMGIVNGHKFMITGEGEGKFPFGTHIILKVKEGGGLPPAYDILTTAFQYGNR 66  
 QY 71 CFSRYPDHMKQHDFFKSAWPEGVQVORTIFFKDDGNVKTAEVKFEGDITLVNRIELKGD 130  
 Db 67 VFTKYPKDIP--DYFKQSPFEGYSWERSMTFEDQGVCTVTSIDIKLEGDCFFYIRFYGVN 124  
 QY 131 FKEDGNILGHK-LEYNYSNHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLADHVOQNTPI 189  
 Db 125 FPSGSGPVMQKTLKWFSTENMVY-----RDGVLLGDVSRLLLEGD----KHHRCNFRS 175  
 QY 190 GDGP--VLLPDNHYLSTQ-SALSKD 211  
 Db 176 TYGAKKGVLPEYFVDHRIEILSHD 201

Search completed: June 21, 2004, 16:00:11  
 Job time : 30.7778 secs

Search completed: June 21, 2004, 16:00:11  
Job time : 30.7778 secs

NRN	[1]
SEQUENCE FROM N.A.	
STRAIN-GFPxm19luv;	
Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,	
Li S.J., Xia N.S.;	
"Colorful mutants of green fluorescent protein from Aequorea	
macrodactyla.";	
Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.	
EMBL; AF435430; AAL33915.1; -;	
GO; GO:0006091; P:energy pathways; IEA.	
InterPro; IPR009017; GFP like.	
InterPro; IPR000786; Green_fl_protein.	
Pfam; PF01353; GFP; 1.	
PRINTS; PR01229; GFLUORESCENT.	
ProDom; PD013756; Green_fl_protein; 1.	
SEQUENCE 238 AA; 27002 MW; BD5BA2982264C018 CRC64;	
Query Match	83.9%; Score 1069; DB 5; Length 238;
Best Local Similarity	81.1%; Pred. No. 1.1e-81;
Matches 193; Conservative	21; Mismatches 24; Indels 0; Gaps 0;
OY	2 VSKGELFTGVPIVLIELDGDVNGHKFVSVEGGDATYGLTKLFICTTGKLPVPWPTL 61
Ddb	1 MSKGELFTGIIVPLVIELDGDVGHGKFVSVRGEEDADYGKLEIKFICTTGTGKLPVPWPTL 60
OY	62 VTTLVGVCGRYPDHMKQHDFFKSAMPEGVQERTIEFKDDGNVKRAEVKFECDTLV 121
OY	61 VTTLISYLICFARYPEHMQNDFFKSAMEGYIQERTIEFDQDGKYKTGEVKFECDTLV 120
OY	122 NRILKIGIDPFKEDGNILGHKLBYNNSHNVYIMADKQNGIKVNFKIRNIEDGSVOLAD 181
Ddb	121 NRILKGMDFPKEDGNILGHKLEYNFNSHNVMIPDKANNGLKVNFKIRNIEGGVOLAD 180
OY	182 HYQQNTPIDGGPVLLPDNHLYLSQSALSDPNKEKRDMVLVGFVTAAGITLGMDELYK 239
Ddb	181 HYQTNPVLGGPVLPIINHYLSQTAISKDRNETRDMVLFEEFFSACGHTHGMDELYK 238
RESULT 14	
QBWTCS	PRELIMINARY; PRT; 238 AA.
ID QBWTCS	
AC QBWTCS;	
DT 01-NAR-2002 (TrEMBLrel. 20, Created)	
DT 01-WAR-2002 (TrEMBLrel. 20, Last sequence update)	
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	
DE Orange fluorescent protein.	
GN GFP.	
OS Aequorea macrodactyla.	
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Leptomedusae;	
OC Aequoreidae; Aequorea.	
OX NCBI_TaxID=147615;	
OX [1]_TaxID=147615;	
RPP SEQUENCE FROM N.A.	
RC STRAIN=OFFxmi;	
RC Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,	
RA Li S.J., Xia N.S.;	
RA "Colorful mutants of green fluorescent protein from Aequorea	
macrodactyla.";	
RT Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.	
RL EMBL; AF435432; AAL33917.1; -;	
DR GO; GO:0006091; P:energy pathways; IEA.	
DR InterPro; IPR009017; GFP like.	
DR InterPro; IPR000786; Green_fl_protein.	
DR Pfam; PF01353; GFP; 1.	
DR PRINTS; PR01229; GFLUORESCENT.	
DR ProDom; PD013756; Green fl protein; 1.	
SQ SEQUENCE 238 AA; 27018 MW; 75521EA5534E573A CRC64;	
Query Match	83.8%; Score 1067; DB 5; Length 238;
Best Local Similarity	81.9%; Pred. No. 1.6e-81;
Matches 195; Conservative	18; Mismatches 25; Indels 0; Gaps 0;
OY	2 VSKGELFTGVPIVLIELDGVNGHKFVSVEGGDATYGLTKLFICTTGKLPVPWPTL 61